THIS WEEK

The permanent revolution

To rediscover its glorious scientific past and build a knowledge-driven economy, Russia must break old habits and loosen state control on research.

Russian science is recovering. After almost two decades of dire financial drought — and despite the casual disdain for all things intellectual shown by the profit-crazed oligarchy who have become Russia's elite — research is reclaiming its place as one of the country's most noble institutions.

Much of the credit for this improved situation must go to Andrei Fursenko, the science and education minister in the government of Prime Minister Vladimir Putin. Fursenko, a physicist trained at the prestigious Ioffe Institute in St Petersburg, understands how modern science works, and knows where and why the Russian research system is in disorder. Not everything he does pleases the Russian academic establishment. But this in itself can be considered an endorsement of Fursenko's approach, given the establishment's inclination to recycle the past rather than turn to modern conventions such as international peer review and scientific competition.

Among the most visible signs of the improved health of science in Russia, and of Fursenko's guiding hand, are the government programmes set up to establish cutting-edge research at Russia's long-neglected universities. These focus in particular on efforts to get experienced Western scientists to do research at Russian university labs through the 'mega-grant' programme, launched last year.

Russia being Russia, Fursenko's efforts have tended to get bogged down by the state's bureaucratic superstructure, to which science and the freedom to pursue it mean very little. As we report on page 17, the most recent example of this is the stalling of a prominent German–Russian mega-grant project to study carbon flux in the environment, which came to a halt on the command of Russia's security services. In this case, Fursenko seems to have won the battle — the project will go ahead, but institutional barriers to collaborative projects remain. Western scientists and companies are learning the hard way that over-regulation in Russia is a different beast to the red tape they encounter at home.

The purchase, import and export of equipment and samples require federal security approval that can be grindingly difficult to obtain. Federal security services need not justify nor explain their rulings. There is no formal way to appeal even obviously offhand decisions, and it is downright impossible for grant holders to communicate with local or federal officers in charge. At lower administrative levels, bribery is yet to be properly addressed, and officials' insistence that every piece of research equipment is purchased through designated Russian agencies (usually at inflated prices) borders on institutional corruption.

Faced with this situation, foreign scientists given mega-grant projects could be forgiven if they elected to do research and spend grant money in their home countries, rather than at the Russian host institutes. This undermines one of the programme's main aims — to bring Russian students and young scientists into contact with high-profile international science early on in their careers — and threatens to diminish its effect on the modernization of Russian science.

Fursenko cannot change the system alone, but must continue to do what he can. All scientists who are participating in the current round of the mega-grant programme, for example, will need clear instructions on deadlines and approval procedures for their projects. And there must be guidance on which formal responsibilities lie with the grant holder, and which ones lie with the host institute.

If Russia is serious in its ambition to develop a knowledge-driven economy, it must substantially reduce the level of state control on research and development. It has given science a helping hand, but — as Fursenko seems to know and as Putin must also understand — further progress needs freedom. ∎

Reality check

Who'd be a scientist? As funding levels fall and competition rises, no one seeking leisure.

The contrast could not be greater. Julie Overbaugh, a lab head at the Fred Hutchinson Cancer Research Center in Seattle, Washington, who researches the molecular virology of HIV, advocates the need for labs that allow their researchers a fulfilling life outside the lab (page 27). Conversely, Alfredo Quiñones-Hinojosa, a stem-cell neurologist and surgeon who heads the brain-tumour programme at Johns Hopkins University in Baltimore, Maryland, drives himself and his lab members as close to a 24/7 working life as is humanly possible (page 20). What might a young scientist make of these two styles, apart from the observation that it takes all sorts?

The necessity for hard work in science has long been emphasized. In his classic *Advice to a Young Scientist*, published in 1979, Peter Medawar emphasized the competitiveness of science and the inevitable concerns about priority. He also issued a golden rule: if you want to make important discoveries, choose an important problem. However, such problems add up to a recipe for perpetual hard work: important problems not only attract the most ambitious scientists but also present risks and false steps in the innovative approaches required to address them.

Overbaugh is right to highlight a need for time away from the bench or computer for creative reflection. Lab heads also need to ensure that their younger lab members maintain a sense of autonomy rather than of cog-in-the-machine. And young scientists applying for posts must understand what sort of lab head they are dealing with. But many older folk wistfully recall their early postdoc careers, when they had one or two clear challenges to focus on late into the night, and over weekends too. As research funding declines in many countries, science will intensify. Anyone lacking the inner intellectual drive and a capacity for relentless focus to get to the heart of the way the world works should stay away. ∎

A. AWAD

The Arab Spring offers hope but no quick fix

Revolutions in Libya and elsewhere have raised hopes for science in the Arab world. But progress will be slow, cautions **Rana Dajani**.

The apparent success of the popular uprising in Libya after months of conflict has once again focused minds on the prospect for fundamental change in the Arab world. Science is not a high priority for countries that have just rid themselves of dictators, but in the wake of the uprisings and protests it is natural for researchers in those nations and colleagues abroad to see opportunities to improve the generally abysmal state of science in Arab countries.

Significant change is unlikely soon. Six months on from the first events of the Arab Spring, there have been no concrete improvements for scientists here in Jordan, and I get the same impression from colleagues in Egypt and Tunisia. The kind of change needed to improve the state of science takes a long time. It is about rebuilding institutions, providing the right environment, obliterating former habits, dismantling bureaucracies, changing mentalities and re-educating people. Such change will take a generation.

One positive thing that I do see and feel is the general attitude of the people, who are more optimistic that things will change for the better. And officials are more reluctant now to exploit and abuse their positions, as they are more likely to be held accountable for their actions. Although the outside world may see headlines about fancy projects such as the building of new institutions, the change to science required in Arab countries is not about bricks-and-mortar improvement but about building intellectual capacity.

There is no lack of minds in the Arab Muslim world, as shown by the many Arab Muslim scientists from these regions doing great science in Western universities. The problem is the environment, which fails to sustain creativity, curiosity and striking out into the unknown — all of which are essential for science to flourish. Such an environment is created only through experience. The whole community must experience the need to find the solutions. And this requires freedom. Dictatorships in Arab Muslim countries thrive on ignorance and fear, and this lack of freedom filters through the whole community, affecting not just political life but reaching down to the household itself — to how parents deal with their children.

I strongly believe that an essential first step towards freeing minds from the habits of the past is to plant the love of reading in our young children. In this way, they revisit other people's experiences across time and space, learn that there are other ways of living, and develop respect for other perspectives. When children read, their horizons expand and they build the confidence to face challenges, create solutions and think without hindrance. I have developed a programme called We Love Reading (www.welovereading.org) that aims to foster a love of reading among children in the

> OFFICIALS ARE MORE **RELUCTANT** NOW TO EXPLOIT AND **ABUSE** THEIR POSITIONS.

Arab world by training women to read aloud to children in their local neighbourhoods.

Through donations and discounts from publishers, we have set up 100 children's libraries across Jordan, and the concept is spreading across the region. By next year, some 20,000 children will have benefited.

Children learn to form their own opinions on the basis of reasoning and deduction. But this requires practice, and this is what we need to encourage. Many students at my university have never formed an independent opinion that reflects their own original thinking. The day I got my students writing essays to express themselves was the day one student told me that he felt human, that he was Someone with a capital S. These are the people who will build our communities and nations, who will make a difference, who will take us into the twenty-first century with confidence and progress.

Alongside this cultural shift, we also need to assess carefully the relationship between Islam and science, particularly in fields with an ethical content, such as stem-cell research. Ethical guidelines for bioengineering and biomedical science for the Muslim world must be drawn up by committees that include scientists, physicians, Islamic scholars and Arabic language specialists.

We have established such a committee at my university, and our discussions indicate that stem-cell research is permissible in Islam, as long as it is carried out with the purpose of improving human health. This conclusion must be re-examined as the field advances.

Such a multidisciplinary approach, new to the Islamic world, is essential to challenge stagnant thinking based on literal interpretations of Islamic sources. The Koran is not a book of scientific facts. It contains verses that describe worldly phenomena, but these are presented as evidence of the elegance and simplicity of creation. Islam is a spiritual guide to life. It teaches us how to live in harmony with ourselves, our fellow humans and the world. There is no conflict between Islam and science.

Islam asks us to use our minds to explore the world around us. It calls for the use of scientific methodology and logic in our approach to science. The verses of the Koran are interpreted by humans, and humans are limited by the scientific knowledge of the era in which the verse was interpreted. The path ahead is not easy, and change will not happen overnight. Still, the Prophet Muhammad said: "Do not belittle any act of good."

In Libya and elsewhere, we can reasonably hope that we have seen acts of good. ∎

Rana Dajani *is assistant professor of molecular biology at the Hashemite University in Zarqa, Jordan.*
e-mail: rdajani@hu.edu.jo

SEVEN DAYS *The news in brief*

Turkish decree

The Turkish government unexpectedly shook up national research politics on 27 August with a decree that gives it the power to nominate the president and vice-presidents of the previously autonomous Turkish Academy of Sciences. The government will also appoint four members of the academy's 14-strong, decision-making council. The decree requires the academy to establish and finance a series of new basic-research institutes, and enables the government to nominate top personnel in TÜBİTAK, the Turkish research-funding agency.

Dollar disclosure

On 23 August, the US National Institutes of Health (NIH) unveiled its new policy for disclosing financial conflicts of interest. In the interests of transparency and public trust, the guidelines impose stricter reporting requirements on both federally funded researchers and their institutions. But the NIH confirmed that it is backing down from an earlier proposal to require disclosure of conflicts of interest on a public website, and is instead allowing institutions to make the information available on request. See go.nature.com/oce4n3 for more.

Greek reform

On 24 August, Greece's parliament passed sweeping reforms to higher education that aim to modernize universities and make it easier for Greek scientists working abroad to return (see *Nature* **475,** 13–14; 2011). The parliament also agreed to abolish a university asylum law intended to stop police intervening in academic affairs. The law prevents police from entering university campuses, but has long been exploited by criminals. It was introduced in 1974 after the fall of the Greek military dictatorship, which had brutally suppressed a student uprising in Athens in 1973.

US energy audit

An audit report from the US Department of Energy's inspector-general has criticized the branch of the department that specializes in funding high-risk, high-pay-off research. The 22 August report says that the Advanced Research Projects Agency-Energy (ARPA-E) lacked policies to ensure oversight and monitoring of awards worth US\$368.6 million, many of them made from funds received through the 2009 stimulus bill. ARPA-E says it is remedying the situation. But the report may fuel political concern: it comes just three months after the House of Representatives recommended cutting the agency's \$180-million budget by 45% for 2012. See go.nature.com/qkojko for more.

Arctic push

Denmark last week announced a ten-year strategy for its priorities in the Arctic, declaring that, together with Greenland and the Faroe Islands, it would welcome industrial development in the region, but also respect the Arctic's fragile environment. Denmark will need to both cooperate and compete with its Scandinavian rivals in Arctic development, as well as with Russia, Canada and the United States.

Ocean drilling

Just two months after the 26-nation Integrated Ocean Drilling Programme (IODP) released a new decadal science plan, the United States has said it cannot afford to take part in the research consortium after 2013. The decision was announced last month by the US National Science Foundation in a letter to the IODP community, and reported in *Science* last

Virginia quake deals seismic surprise

A magnitude 5.8 earthquake in rural Virginia on 23 August caused disruption across broad swathes of the eastern United States, where it was the most significant tremor in a century. Major earthquakes are rare in the region because its crust is old and mostly stable, but when quakes do occur, the strong crustal rock transmits seismic waves with relatively little loss of energy, so they can cover vast distances. Shaking was felt along the eastern seaboard, from Florida to Nova Scotia in Canada; and the quake caused an automatic shutdown at the nearby North Anna nuclear power plant in Mineral, Virginia. There was some damage to buildings in urban centres, including Washington DC (its cathedral is pictured). See go.nature.com/sgugvi for more.

week. Yet it was not entirely unexpected — worries had been building since the spring, when it emerged that rising fuel costs were limiting the activities of the US drilling ship *JOIDES Resolution* (see *Nature* **473,** 137; 2011). IODP partners are now discussing whether a less-costly cooperation might allow US scientists to take part in international ocean-drilling projects after 2013.

RESEARCH

Exploding star

The closest type Ia supernova in nearly 40 years has been spotted in the spiral galaxy M101 by astronomers at the Palomar Observatory in California. As the brightest and most energetic kind of stellar explosion, it can be used to measure the accelerating expansion of the Universe. Astronomers are now scrambling to study the event as it brightens over the coming weeks. See go.nature.com/oraf5x for more.

Ethics report

US Public Health Service researchers knew that they were acting unethically when they exposed hundreds of Guatemalans to sexually transmitted diseases in the 1940s without the subjects' consent, according to the Presidential Commission for the Study of Bioethical Issues.

The commission will now assess whether current research rules protect trial participants from similar abuses, and report its conclusions in December.

EVENTS

Storm damage

The hurricane season has definitely arrived. Torrential rain, flash floods and 140 km h^{-1} winds left a trail of destruction along the east coast of the United States this week. Irene (**pictured**) may have been downgraded from hurricane status to a category 1 storm by the time it made landfall in North Carolina, but it still left millions without power and caused more than 40 deaths as it moved northwards. Irene was the ninth named storm of the season; forecasters predict up to 19 this year.

Space failure

The next manned mission to the International Space Station will be delayed by a month,

to October, following the loss of a Russian cargo capsule carrying fresh supplies into orbit. The launch vehicle, a Soyuz-U rocket, was lost after its third stage failed around 5 minutes after lift-off from the Baikonur Cosmodrome in Kazakhstan. The failure raises questions about quality control in the Russian space programme: in August, a Russian telecommunications satellite aboard a Proton rocket was delivered to the wrong orbit. See go.nature.com/lfp5zl for more.

PEOPLE

Research head

Donald Dingwell is the new secretary-general of the European Research Council (ERC), Europe's competitive funding agency. The geoscientist, currently at Ludwig Maximilian University in Munich, Germany, will act as a liaison between the ERC and the European Commission. He fills a year-long vacancy left after the departure of the previous incumbent, Andreu Mas-Colell.

Iran assassin

The alleged assassin of an Iranian particle physicist killed by a bomb explosion last year pleaded guilty in a Tehran court on 24 August. Majid Jamali Fashi confessed

to setting up the bomb that on 12 January 2010 killed Masoud Alimohammadi on his way to work at the University of Tehran (see *Nature* **463,** 279; 2010). Prosecutors said that Israel's intelligence agency Mossad was ultimately behind the murder — but observers worry that the confession was a show trial. Another Iranian physicist was killed last November, and an Iranian electrical engineering student was murdered in July. See go.nature.com/tukvtk for more.

Researcher returns

Wildlife biologist Charles Monnett of the US Bureau of Ocean Energy Management, Regulation and Enforcement (BOEMRE) went back to work on 26 August after a six-week suspension triggered by an investigation by the US Department of the Interior's inspector-general. BOEMRE says that Monnett, who in 2006 first observed polar bears that had apparently drowned while searching for sea ice, remains under investigation. He will now do environmental-assessment research rather than return to his former role managing research contracts. See go.nature.com/ebvtvb for more.

↻ **NATURE.COM**
For daily news updates see:
www.nature.com/news

TREND WATCH

Surveys in the United Kingdom and the United States suggest that the rate of increase in obesity is slowing. Even so, the prevalence of obesity in US adults is set to grow from around 32% in 2007–08 to 50% in 2030 for men, and from 35% to 45% for women. The projections were published on 27 August (Y. C. Wang *et al. Lancet* **378,** 815–825; 2011). The study also estimates that by 2030, the annual cost of treating obesity-related diseases in the United States will have risen by US$48 billion–66 billion.

THE OBESITY TIDE

Projections suggest that more than 50% of US male adults are likely to be obese by 2030.
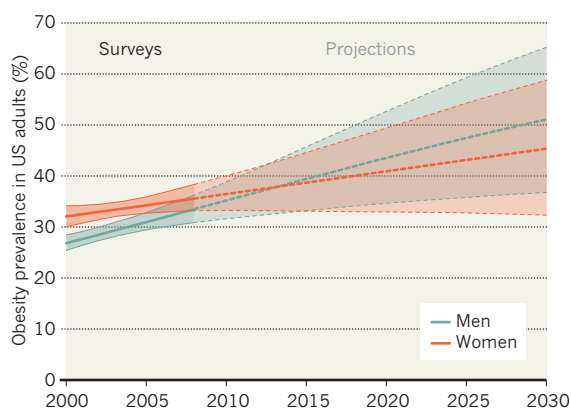
NEWS IN FOCUS

Yoshihiko Noda is Japan's sixth prime minister in five years, replacing Naoto Kan after his resignation.

ENERGY

Japan's new leader faces energy gap

Power policy remains in flux.

BY DAVID CYRANOSKI IN TOKYO

Yoshihiko Noda, Japan's new prime minister, has a daunting in-tray to deal with: a faltering economy, a huge reconstruction effort following the devastating earthquake and tsunami in March, and an ongoing nuclear emergency at the Fukushima Daiichi power plant. Elected on 29 August after Naoto Kan's resignation last week as leader of the ruling Democratic Party of Japan,

Noda must also contend with the threat of a complete vacuum where the country's energy policy should be.

After the Fukushima disaster, Kan vowed to "leave nuclear energy behind", but made no clear plans to fill the energy gap. And neither Noda nor the other leadership contenders articulated a clear position on energy during the brief election campaign. When pressed on the issue, they steered clear of strong statements about the future of nuclear power or how they would

keep the lights on in its absence. "We have no energy policy," says Tatsuo Oyama, who studies investment in power utilities at the National Graduate Institute for Policy Studies in Tokyo. "It's a serious issue that Noda has to deal with."

Last year, Japan committed to constructing 14 new reactors so that nuclear power would provide half of the country's electricity by 2030. But after the accident at Fukushima, public mistrust of the technology prompted Kan to ditch the plan. He also ordered existing reactors to be taken offline for a series of stress tests, in addition to normal inspections. All but one of the country's 54 existing reactors are set to be shut down by May 2012, removing a quarter of Japan's power capacity. Restarting them would require approval from local governments, which they may be reluctant to grant.

Kan had promised that renewables and efficiency measures would make up the shortfall, but did not explicitly say how. "He made a change but he did it without debate and without a road map," says Tsutomo Toichi, an energy specialist at the Japanese Institute of Energy Economics in Tokyo. "It was very abrupt."

The short-term consequence has been a jump in the use of fossil fuels, including an estimated 20% increase in imports of expensive liquefied natural gas. Toichi says that there could be a 20% increase in the cost of producing electricity in 2012 if nuclear reactors are shut down as expected, probably leading to higher rates for consumers. This week, the Tokyo Electric Power Company announced that it is considering a 10% hike in next year's electricity rate. Citizens' willingness to economize has trimmed power demand in Japan, but Oyama believes that the sagging economy has also reduced electricity consumption. Hiromasa Yonekura — chairman of Sumitomo Chemical, headquartered in Tokyo and Osaka, and head of the powerful Japan Business Federation (Keidanren) — has repeatedly warned that high prices could force companies to move operations overseas.

To prevent this, Toichi says that the new government must get the nuclear plants back online as soon as possible. Despite public opposition, Noda has hinted that nuclear will continue to make up part of Japan's energy ▶

H. YAMAGUCHI/CORBIS

mix, and industry minister Banri Kaieda announced for the first time this week that he expects some reactors to restart this year.

That would be a step in the wrong direction, says Tetsunari Iida, director of the Tokyo-based Institute for Sustainable Energy Policies, who wants the country to seize the opportunity to invest in renewable energy.

On 12 September, Iida will launch the Japan Renewable Energy Foundation, which is backed by ¥1 billion (US$13 million) from Japan's richest man, telecoms mogul Masayoshi Son. The foundation will bring together some 100 experts from around the world to analyse obstacles to implementing renewable energy, and offer policy recommendations to the new government.

The foundation's cause did get a boost from Kan, who had agreed to step down only if Japan's parliament passed a bill to support clean energy. The bill, which passed on 26 August, will next year guarantee a minimum price for wind, solar and other renewable energies that will make them more attractive for suppliers to invest in.

Iida is confident that this will help to sustain the shift away from conventional power sources. He notes that, from 2008 to 2010, Japan's annual increase in solar-power capacity jumped from 230 megawatts to almost 1 gigawatt, giving the country a total capacity of 3.6 gigawatts. He believes that the renewables bill will make huge solar farms economically attractive for the first time. "I think next year we'll see a tenfold jump in solar and fivefold in wind," he predicts. Many say that this is unrealistic. "It's impossible within two years, and too ambitious within ten," says Toichi.

A bill on global warming could also boost renewables by preventing Japan from filling its energy gap with fossil fuels. But the bill's prospects are dim, even though its aim — a legally binding target to reduce greenhouse-gas emissions to 25% below 1990 levels — was once a cornerstone of the Democratic Party's manifesto. "Because of the reduction in our reliance on nuclear power, it will be very difficult" to hit that target, says Toichi.

> **"There is now a broad consensus that we need to reduce our reliance on nuclear power."**

Japan's parliament has so far declined to consider the bill, and political support for it is waning rapidly. "The new majority position [within the Democratic Party] is to get rid of the 25% target," says Iida.

Yet despite the uncertainties that now cloud Japan's future, Kan's administration did at least deliver one clear achievement on energy policy, says Iida. "There is now a broad consensus that we need to reduce our reliance on nuclear power. That is the atmosphere now." ∎

Variation among *Arabidopsis* strains can reflect genetic adaptation to their local habitat.

D. WEIGEL

GENOMICS

Halfway point for 1,001 genomes quest

Plant researchers map diversity of Arabidopsis thaliana.

BY HEIDI LEDFORD

Joseph Ecker's teenage son listened intently as his father told him about the 1000 Genomes Project, which aims to sequence and compare the genomes of 1,000 people. Ecker, a molecular geneticist, explained that he and his colleagues were launching a similar project for the plant *Arabidopsis thaliana*. "My son said, 'Well then you should sequence 1,001,'" Ecker recalls. "He's a very competitive kid."

And so the *Arabidopsis* '1001 Genome Project' was born. More than four years later, a loose confederation of laboratories is on the verge of making that challenge a reality. Papers published online in *Nature*[1] and *Nature Genetics*[2] this week report the sequencing of nearly 100 *A. thaliana* genomes, the first swathe released by the project; around 400 more have been sequenced, but are not yet ready for publication. Last week, Ecker's group at the Salk Institute in La Jolla, California, won a US$2-million grant from the National Science Foundation (NSF) to polish off another 500 strains, and to catalogue expressed RNAs and map DNA methylation, a chemical modification that affects gene expression.

Arabidopsis thaliana, or thale cress, is a small weed with a simple genome that stands in as a genetic reference for plants with more complex genomes. The genome project aims to uncover genetic changes that enable plants to adapt to their local environments. There are thousands of strains of *A. thaliana* in stocks worldwide, each of which might carry unique traits that helped it to thrive in its natural environment — tolerance for drought, perhaps, or defences against viral pathogens. "If you learn which genes are important for these traits, you could breed them into crops — to allow them to move into a new environment or continue to succeed where they face climate change," Ecker says.

The mining of natural variation for genetic information has gained momentum as faster DNA sequencing has delivered multiple genomes from wild populations. Similar projects are under way in mice, fruitflies, rice and, of course, humans. "If you go into nature, you find all these fascinating mutations that have survived the sieve of natural selection," says geneticist Trudy Mackay of North Carolina State University in Raleigh, who leads the work in fruitflies. "But in the past we've been hampered in our ability to tease them apart."

The 1001 Genome Project has had some problems, however. Unable to get funding for a single project, participating labs went their own ways, getting grants from a variety of sources, says Detlef Weigel, a plant biologist at the Max Planck Institute for Developmental Biology in Tübingen, Germany, who has spearheaded the project. The result was a fragmented effort, with each group sequencing strains and using techniques that best fitted its own research.

And Ecker frets that this ad hoc coalition won't even have a central place to deposit and organize its data. *Arabidopsis* researchers have relied on The Arabidopsis Information Resource (TAIR), but NSF funding for that project is being phased out and its fate is unclear. "We don't want to have these data scattered all over the place," says Ecker, "but there may be nowhere to put them." ∎

1. Gan, X. *et al. Nature* http://dx.doi.org/10.1038/nature10414 (2011).
2. Cao, J. *et al. Nature Genet.* http://dx.doi.org/10.1038/ng.911 (2011).

▶ mix, and industry minister Banri Kaieda announced for the first time this week that he expects some reactors to restart this year.

That would be a step in the wrong direction, says Tetsunari Iida, director of the Tokyo-based Institute for Sustainable Energy Policies, who wants the country to seize the opportunity to invest in renewable energy.

On 12 September, Iida will launch the Japan Renewable Energy Foundation, which is backed by ¥1 billion (US$13 million) from Japan's richest man, telecoms mogul Masayoshi Son. The foundation will bring together some 100 experts from around the world to analyse obstacles to implementing renewable energy, and offer policy recommendations to the new government.

The foundation's cause did get a boost from Kan, who had agreed to step down only if Japan's parliament passed a bill to support clean energy. The bill, which passed on 26 August, will next year guarantee a minimum price for wind, solar and other renewable energies that will make them more attractive for suppliers to invest in.

Iida is confident that this will help to sustain the shift away from conventional power sources. He notes that, from 2008 to 2010, Japan's annual increase in solar-power capacity jumped from 230 megawatts to almost 1 gigawatt, giving the country a total capacity of 3.6 gigawatts. He believes that the renewables bill will make huge solar farms economically attractive for the first time. "I think next year we'll see a tenfold jump in solar and fivefold in wind," he predicts. Many say that this is unrealistic. "It's impossible within two years, and too ambitious within ten," says Toichi.

A bill on global warming could also boost renewables by preventing Japan from filling its energy gap with fossil fuels. But the bill's prospects are dim, even though its aim — a legally binding target to reduce greenhouse-gas emissions to 25% below 1990 levels — was once a cornerstone of the Democratic Party's manifesto. "Because of the reduction in our reliance on nuclear power, it will be very difficult" to hit that target, says Toichi.

"There is now a broad consensus that we need to reduce our reliance on nuclear power."

Japan's parliament has so far declined to consider the bill, and political support for it is waning rapidly. "The new majority position [within the Democratic Party] is to get rid of the 25% target," says Iida.

Yet despite the uncertainties that now cloud Japan's future, Kan's administration did at least deliver one clear achievement on energy policy, says Iida. "There is now a broad consensus that we need to reduce our reliance on nuclear power. That is the atmosphere now." ∎

Variation among *Arabidopsis* strains can reflect genetic adaptation to their local habitat.

GENOMICS

Halfway point for 1,001 genomes quest

Plant researchers map diversity of Arabidopsis thaliana.

BY HEIDI LEDFORD

Joseph Ecker's teenage son listened intently as his father told him about the 1000 Genomes Project, which aims to sequence and compare the genomes of 1,000 people. Ecker, a molecular geneticist, explained that he and his colleagues were launching a similar project for the plant *Arabidopsis thaliana*. "My son said, 'Well then you should sequence 1,001,'" Ecker recalls. "He's a very competitive kid."

And so the *Arabidopsis* '1001 Genome Project' was born. More than four years later, a loose confederation of laboratories is on the verge of making that challenge a reality. Papers published online in *Nature*[1] and *Nature Genetics*[2] this week report the sequencing of nearly 100 *A. thaliana* genomes, the first swathe released by the project; around 400 more have been sequenced, but are not yet ready for publication. Last week, Ecker's group at the Salk Institute in La Jolla, California, won a US$2-million grant from the National Science Foundation (NSF) to polish off another 500 strains, and to catalogue expressed RNAs and map DNA methylation, a chemical modification that affects gene expression.

Arabidopsis thaliana, or thale cress, is a small weed with a simple genome that stands in as a genetic reference for plants with more complex genomes. The genome project aims to uncover genetic changes that enable plants to adapt to their local environments. There are thousands of strains of *A. thaliana* in stocks worldwide, each of which might carry unique traits that helped it to thrive in its natural environment — tolerance for drought, perhaps, or defences against viral pathogens. "If you learn which genes are important for these traits, you could breed them into crops — to allow them to move into a new environment or continue to succeed where they face climate change," Ecker says.

The mining of natural variation for genetic information has gained momentum as faster DNA sequencing has delivered multiple genomes from wild populations. Similar projects are under way in mice, fruitflies, rice and, of course, humans. "If you go into nature, you find all these fascinating mutations that have survived the sieve of natural selection," says geneticist Trudy Mackay of North Carolina State University in Raleigh, who leads the work in fruitflies. "But in the past we've been hampered in our ability to tease them apart."

The 1001 Genome Project has had some problems, however. Unable to get funding for a single project, participating labs went their own ways, getting grants from a variety of sources, says Detlef Weigel, a plant biologist at the Max Planck Institute for Developmental Biology in Tübingen, Germany, who has spearheaded the project. The result was a fragmented effort, with each group sequencing strains and using techniques that best fitted its own research.

And Ecker frets that this ad hoc coalition won't even have a central place to deposit and organize its data. *Arabidopsis* researchers have relied on The Arabidopsis Information Resource (TAIR), but NSF funding for that project is being phased out and its fate is unclear. "We don't want to have these data scattered all over the place," says Ecker, "but there may be nowhere to put them." ∎

1. Gan, X. *et al. Nature* http://dx.doi.org/10.1038/nature10414 (2011).
2. Cao, J. *et al. Nature Genet.* http://dx.doi.org/10.1038/ng.911 (2011).

NUCLEAR PHYSICS

Isotope factory accelerates

US nuclear scientists find silver lining in economic downturn — lower construction costs.

BY EUGENIE SAMUEL REICH

Given the tough economic times, good news was the last thing that nuclear scientists expected at an 18–20 August meeting for potential users of the Facility for Rare Isotope Beams (FRIB), a planned national facility for nuclear physicists, which will be run under the auspices of the US Department of Energy (DOE). Instead, they heard that the downturn may have an upside: construction of the US$614.5-million facility, originally slated to begin in 2013, could be brought forward a year. Cash-strapped Michigan State University in East Lansing, where FRIB is based, has decided to pump in money while construction costs remain low.

"I'm delighted," says Ani Aprahamian, a physicist at the University of Notre Dame in Indiana who hopes to use FRIB to study short-lived isotopes that are key to the production of heavy elements in stars. "I have students whose futures depend on these studies, and it would be great to be able to do them sooner."

FRIB, expected to serve around 800 users a year, will accelerate ionized atoms down a 500-metre-long series of tunnels folded around like a paper clip and then shatter them against a graphite target to produce beams of rare isotopes at higher intensity than at any other facility in the world. The fragments could include thousands of isotopes that are predicted but have never been seen on Earth (see 'Probing instability').

US nuclear scientists have dreamt of such a facility since the late 1980s, hoping that studying the lifetimes, masses, excited states and structure of rare isotopes might shed light on fundamental questions in nuclear physics and astrophysics. A report from the National Academy of Sciences endorsed the idea in 2007, and in 2009 the DOE's Office of Science signed a cooperative agreement with Michigan State University to build the accelerator by 2020, with a $520-million commitment from the DOE and another $94.5 million from the university.

The proposal to accelerate construction and be ready to begin doing science as early as 2018 is largely the brainchild of physicist and FRIB's project manager Thomas Glasmacher, who says that while thinking of ways to protect the project he came up with the idea of locking in construction costs at low prices. "It's a really good time to build things," says Glasmacher. "People are out of work and that's driving prices down." Calculating that once the US economy rebounded, construction costs would rise fast, he pitched the case to the university. In February the university president agreed, even though a Michigan state budget proposed the same month would cut $69 million in state support for the university. Glasmacher notes that he didn't ask for any extra money. The university will simply allocate $15 million of the promised budget a year earlier than planned. The next challenge for Glasmacher will be persuading the DOE to agree to the accelerated schedule, which he will propose at a peer-review meeting later this month.

Moving FRIB ahead would help remedy the loss of the Holifield Radioactive Ion Beam Facility at Oak Ridge National Laboratory in Tennessee, which the DOE has announced it will stop funding (see *Nature* **471**, 278; 2011), a move to save $10.3 million per year. "The Holifield community supports FRIB," says Witold Nazarewicz, scientific director of the Holifield facility. Aprahamian adds that speeding up FRIB will also enable US nuclear science to compete sooner with facilities already running or planned overseas. The higher intensity of FRIB's beams should enable experimentalists to get reliable statistics on hundreds of isotopes at once, compared with a handful at other facilities.

*"The quicker you do a project, the cheaper it is."*

FRIB is not the only science facility to have leveraged low construction costs. "We were fortunate to have negotiated our major contracts at a time when the construction industry was hit hard by the economic downturn," says Steve Dierker, director of the $912-million National Synchrotron Light Source II at Brookhaven National Laboratory in Upton, New York, a future source of high-energy X-rays that is around 60% complete.

However, FRIB's director, Konrad Gelbke, cautions that the whole project still depends on the availability of funding through Congress, and on clearing multiple hurdles at the DOE. A statement from the Office of Science says that a review is scheduled for the spring of 2012 to assess FRIB's readiness to start construction, and that if the review is favourable, the DOE will consider approving construction. "DOE appreciates the continuing willingness of Michigan State University to be flexible in apportioning its part of the cost-share," the statement says.

Glasmacher is confident that his approach is the right one and that the DOE reviewers will see that. "We're presenting it as an opportunity, because the quicker you do a project the cheaper it is," he says. ∎

PROBING INSTABILITY

The planned Facility for Rare Isotope Beams (FRIB) will generate isotopes that have predicted but previously undetected ratios of protons to neutrons.
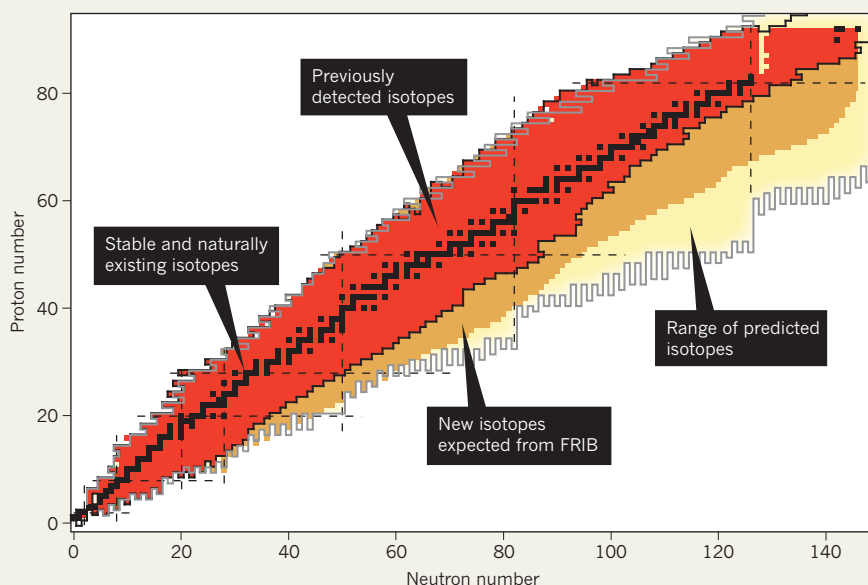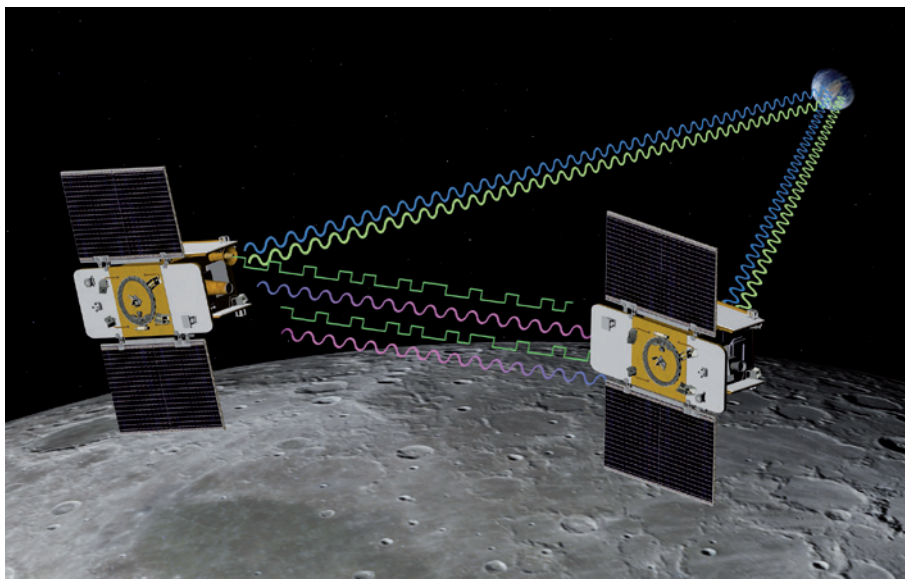
By working together, GRAIL's two probes will map the gravity of both the near and far sides of the Moon.

LUNAR SCEINCE

Twins to probe Moon's heart

NASA mission will survey lunar gravity to map the dense rock beneath the surface.

BY ERIC HAND

The Moon's face is an open book, but its deeper nature is still a mystery. Probes from Europe, Japan, India, China and the United States have imaged the lunar surface in exquisite detail, mapped its minerals, looked for evidence of water and scouted for potential landing sites. Now Maria Zuber, principal investigator for NASA's Gravity Recovery and Interior Laboratory (GRAIL) mission, which is set to launch on 8 September, wants to reveal the Moon's hidden history.

"I think we're going to find something quite surprising," says Zuber, a geophysicist at the Massachusetts Institute of Technology in Cambridge. Her confidence stems from the maps of lunar gravity that GRAIL is set to provide — orders of magnitude better than any before. The density variations they aim to reveal in the subsurface rock should shed light on the Moon's tumultuous and geologically active past, help to determine whether it has a liquid core and yield clues to the underlying structure of its giant impact basins, the lunar 'maria'.

The GRAIL mission consists of twin spacecraft that are near replicas of the Gravity Recovery and Climate Experiment (GRACE),

a pair of satellites that have orbited Earth since 2002, mapping the planet's gravity field so finely that they could see shifts in groundwater aquifers and ocean currents. Adapting that flight-tested technology helped to keep the cost of GRAIL to US$496 million. Another saving came in mission design: instead of blasting straight to the Moon, the spacecraft will ease into a lunar polar orbit after a three-and-a-half month journey, and only a modest amount of fuel will be needed for slowing down.

Apart from four cameras on each spacecraft, which will capture images for public outreach, GRAIL carries only one instrument — and it isn't even pointed at the Moon. As the two spacecraft coast 55 kilometres above the lunar surface, and 60–225 kilometres apart, a high-frequency radio link will measure the exact distance between them. As one probe approaches a high-density object — for example, a mountain — it will feel a slightly stronger gravitational force and will momentarily speed up, changing the separation from its companion. For GRAIL to deliver on its promise, such movements will need to be measured to a fraction of a micrometre. Such precision

will require taking into account the gravitational influence of distant planets, the movement of tectonic plates under tracking stations on Earth and even the pressure of sunlight on the spacecrafts' solar panels, says Zuber.

The Moon's gravity has already been mapped less precisely, by measuring, from Earth, the shifts in the speed of a single lunar orbiter — but this is impossible when the orbiter goes behind the Moon. SELENE (Kaguya), a Japanese mission launched in 2007, mapped gravity on the far side with the help of a relay satellite that orbited at a higher altitude, within radio-sight of both Earth and SELENE. But Zuber says that GRAIL's maps will be far more precise — even better than GRACE's maps of Earth, because the effects of Earth's atmospheric drag mean that the GRACE spacecraft have to orbit at altitudes ten times higher than GRAIL will.

One science target will be the Moon's deep structure. By bouncing lasers off reflectors left by the Apollo astronauts, researchers have picked up hints of a subtle wobbling, suggesting the presence of a soft core. Zuber says GRAIL should be able to confirm those hints, and might also find surprising additions to the core — for instance, by discerning whether compounds such as titanium oxide crystallized and sank into the core when the Moon was initially a ball of magma.
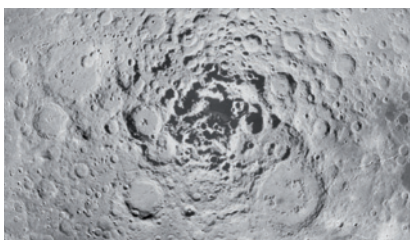
The overall findings could illuminate how planets in the inner Solar System cooled into layered structures. "It transcends just knowing about the Moon," says geologist Brad Jolliff of Washington University in St Louis, Missouri. "It helps us to understand how other rocky bodies differentiate." By probing the rocks around impact basins, Zuber says, GRAIL should also help modellers to understand the dynamics of giant impacts.

NASA's plans for exploring the Moon after GRAIL's 90-day mission are uncertain, says Chip Shearer, a planetary geologist at the University of New Mexico in Albuquerque and chair of NASA's Lunar Exploration Analysis Group. A wave of missions designed to support NASA's now-defunct Constellation programme, which envisioned returning humans to the Moon by 2020, is tapering off (see 'Moon rush'). The last of these, the Lunar Atmosphere and Dust Environment Explorer (LADEE), a mission to measure the effects of the Moon's fine dust, is due to launch in 2013. But two lunar science priorities aren't being addressed. A proposed mission, led by Jolliff, to return samples from the Moon's largest impact basin, near the south pole, lost out to a mission to return an asteroid sample. And advocates for the International Lunar Network, once a NASA-led endeavour to set up seismic detectors on the Moon's surface, have been told that they will have to compete for NASA's support with everybody else's Solar System proposals.

"We know what, in terms of lunar science, should come next," says Shearer. "But we don't know what will come next." ∎

NATURE.COM
Read about mapping Earth's gravity at:
go.nature.com/yp1znv

MOON RUSH

The wealth of lunar missions in the past two decades is not set to last.

1994

NASA's Clementine hints at existence of water at lunar south pole (pictured).

1998

NASA's Lunar Prospector allows gravity map of the Moon's near side.

2004

United States unveils 'vision' to return astronauts to the Moon.

SMART-1, Europe's first lunar orbiter, launched.

2007

Japan's SELENE (Kaguya) produces gravity map of the far side.

Launch of Chang'e, China's first lunar orbiter.

Google supports US$30-million Lunar X Prize for first private team to put a rover on the Moon.

2008

India's Chandrayaan-1 finds evidence of watery minerals.

2009

NASA's Lunar Reconnaissance Orbiter finds water in shadowed polar craters.

2010

Chang'e 2 launches.

NASA human space policy revised to target asteroids before the Moon.

2011

GRAIL due to launch.

2013

NASA's Lunar Atmosphere and Dust Environment Explorer (pictured, below) is scheduled to launch.

COLLABORATION

Red tape puts chill on Siberian research

High-profile carbon project to proceed, but with a proviso.

BY QUIRIN SCHIERMEIER

One of Russia's most prominent international science projects has fallen foul of cold-war-era concerns. An expedition to study carbon transport around Siberia's Yenisey River has been postponed for a year after officials blocked the use of sampling equipment and put some sites off-limits. The episode highlights the tension between Russia's bureaucracy and its growing ambition to develop wider research collaborations — part of a strategy to revitalize domestic science.

In July, Ernst-Detlef Schulze, a carbon-cycle researcher and founding director of the Max Planck Institute for Biogeochemistry in Jena, Germany, and 35 scientists from Germany, the Netherlands, France and Russia, had their bags packed for the journey to Siberia.

Then they received a letter from the Russian Federal Security Service prohibiting them from using any Western equipment on the trip. Schulze, who last year received a 150-million-rouble (US$5-million) grant from the Russian government to do research in Siberia, was aghast. "When I first saw the letter I just couldn't believe what I read," he says. "I was so disappointed and furious I went to my own forest and lumbered trees until I was exhausted."

Western scientists participating in the 12-billion-rouble 'mega-grant' programme that funds Schulze (see *Nature* **473**, 428–429; 2011) have previously complained about excessive bureaucracy and frequent problems with export control and customs, and have received support from the Russian science minister, Andrei Fursenko (see *Nature* **465**, 858; 2010).

So Schulze and his co-workers at the Siberian Federal University in Krasnoyarsk used diplomatic channels to try to reverse the ruling. Equipment worth €600,000 (US$865,000) was already in Siberia, ready to be used to calculate the carbon budget of the Yenisey's huge catchment area (see map). Siberia's soils and forests serve as one of Earth's largest carbon sinks, but the carbon flux between ecosystems there has never been studied in detail.

At a hastily arranged meeting in Moscow on 4 August, Fursenko said that he regretted the situation and promised to try to find a solution. A week later, the expedition was given clearance to begin on 1 September — but with restrictions that Schulze found unacceptable. Sampling tools, for instance, could be handled only by

SIBERIAN SAMPLING
Researchers will study the flow of carbon between land, water and air across various climate zones in Siberia.

D. SCHULZE

Russian staff — most of whom are not trained to use the geochemical analysis equipment, says Schulze. The Russian Federal Services for Technology and Export Control also refused to approve the purchase of equipment in the West, such as a freeze-dryer to preserve samples.

"This is not the way to do science," says Han Dolman, an environmental scientist at the VU University Amsterdam. Dolman had planned to study how carbon is transported from soils into the Yenisey River, and how much of the organic carbon dissolved in the river becomes carbon dioxide. Unlike on land, carbon exchange in aquatic systems is poorly understood.
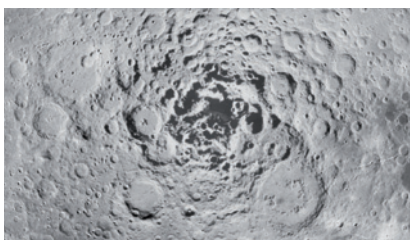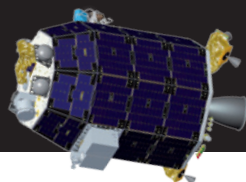
On 12 August, Schulze threatened to cancel the expedition, but further negotiations yielded a compromise: the trip will take place next year, with the only restriction being a ban on taking samples from a roughly 20,000-square-kilometre area around Krasnoyarsk, where Russia operates a nuclear-reprocessing facility.

The saga shows that scientists do not yet have enough freedom in their research in Russia, says Schulze. "There's an absurdly opaque and often arbitrary bureaucracy at work. Thankfully, Fursenko was always clearly on our side." ∎

SEE EDITORIAL P.5

MOON RUSH

The wealth of lunar missions in the past two decades is not set to last.

1994

NASA's Clementine hints at existence of water at lunar south pole (pictured).

1998

NASA's Lunar Prospector allows gravity map of the Moon's near side.

2004

United States unveils 'vision' to return astronauts to the Moon.

SMART-1, Europe's first lunar orbiter, launched.

2007

Japan's SELENE (Kaguya) produces gravity map of the far side.

Launch of Chang'e, China's first lunar orbiter.

Google supports US$30-million Lunar X Prize for first private team to put a rover on the Moon.

2008

India's Chandrayaan-1 finds evidence of watery minerals.

2009

NASA's Lunar Reconnaissance Orbiter finds water in shadowed polar craters.

2010

Chang'e 2 launches.

NASA human space policy revised to target asteroids before the Moon.

2011

GRAIL due to launch.

2013

NASA's Lunar Atmosphere and Dust Environment Explorer (pictured, below) is scheduled to launch.

NASA

NASA/AMES

COLLABORATION

Red tape puts chill on Siberian research

High-profile carbon project to proceed, but with a proviso.

BY QUIRIN SCHIERMEIER

One of Russia's most prominent international science projects has fallen foul of cold-war-era concerns. An expedition to study carbon transport around Siberia's Yenisey River has been postponed for a year after officials blocked the use of sampling equipment and put some sites off-limits. The episode highlights the tension between Russia's bureaucracy and its growing ambition to develop wider research collaborations — part of a strategy to revitalize domestic science.

In July, Ernst-Detlef Schulze, a carbon-cycle researcher and founding director of the Max Planck Institute for Biogeochemistry in Jena, Germany, and 35 scientists from Germany, the Netherlands, France and Russia, had their bags packed for the journey to Siberia.

Then they received a letter from the Russian Federal Security Service prohibiting them from using any Western equipment on the trip. Schulze, who last year received a 150-million-rouble (US$5-million) grant from the Russian government to do research in Siberia, was aghast. "When I first saw the letter I just couldn't believe what I read," he says. "I was so disappointed and furious I went to my own forest and lumbered trees until I was exhausted."

Western scientists participating in the 12-billion-rouble 'mega-grant' programme that funds Schulze (see *Nature* **473,** 428–429; 2011) have previously complained about excessive bureaucracy and frequent problems with export control and customs, and have received support from the Russian science minister, Andrei Fursenko (see *Nature* **465,** 858; 2010).

So Schulze and his co-workers at the Siberian Federal University in Krasnoyarsk used diplomatic channels to try to reverse the ruling. Equipment worth €600,000 (US$865,000) was already in Siberia, ready to be used to calculate the carbon budget of the Yenisey's huge catchment area (see map). Siberia's soils and forests serve as one of Earth's largest carbon sinks, but the carbon flux between ecosystems there has never been studied in detail.

At a hastily arranged meeting in Moscow on 4 August, Fursenko said that he regretted the situation and promised to try to find a solution. A week later, the expedition was given clearance to begin on 1 September — but with restrictions that Schulze found unacceptable. Sampling tools, for instance, could be handled only by

SIBERIAN SAMPLING
Researchers will study the flow of carbon between land, water and air across various climate zones in Siberia.

D. SCHULZE

Russian staff — most of whom are not trained to use the geochemical analysis equipment, says Schulze. The Russian Federal Services for Technology and Export Control also refused to approve the purchase of equipment in the West, such as a freeze-dryer to preserve samples.

"This is not the way to do science," says Han Dolman, an environmental scientist at the VU University Amsterdam. Dolman had planned to study how carbon is transported from soils into the Yenisey River, and how much of the organic carbon dissolved in the river becomes carbon dioxide. Unlike on land, carbon exchange in aquatic systems is poorly understood.

On 12 August, Schulze threatened to cancel the expedition, but further negotiations yielded a compromise: the trip will take place next year, with the only restriction being a ban on taking samples from a roughly 20,000-square-kilometre area around Krasnoyarsk, where Russia operates a nuclear-reprocessing facility.

The saga shows that scientists do not yet have enough freedom in their research in Russia, says Schulze. "There's an absurdly opaque and often arbitrary bureaucracy at work. Thankfully, Fursenko was always clearly on our side." ■

SEE EDITORIAL P.5

POLICY

Scientists promised 'one voice' in European policy

ScienceEurope hopes to shift balance of power away from Brussels and towards researchers.

BY NATASHA GILBERT

Scientists often struggle to get their opinions heard above the din of voices competing to influence policy and research-funding decisions in the European Union. A new Brussels-based group, ScienceEurope, is now positioning itself as the scientists' champion in the fight to sway decision-makers. "We will become the single voice for science in Europe," says Paul Boyle, a member of the pilot board of the organization, which launches next month.

"If we speak in one voice it will be easier to see if our recommendations have influenced policy," adds Boyle, who is chief executive of Britain's Economic and Social Research Council in Swindon.

ScienceEurope unites two science advocacy groups: the European Science Foundation (ESF) based in Strasbourg, France, and the European Heads of Research Councils (EURO-HORCs) based in Berne. The two groups have common members, including Europe's leading national research and funding organizations, such as the Helmholtz Association of German Research Centres, headquartered in Berlin, and Britain's Medical Research Council in London. They also share similar goals and have previously worked together on policy development. Earlier this year they voted to join forces.

Once ScienceEurope holds its founding assembly in Berlin on 21 October, EURO-HORCs will cease to exist. The ESF may continue as a separate body but will probably wind down its activities over the next few years.

Marja Makarow, a molecular biologist at the University of Helsinki and chief executive of the ESF, says that the merger was born out of a need for an organization that would have greater influence in Brussels and would be "flexible

WHO FUNDS EUROPE'S SCIENCE?
Despite having a major influence on research agendas, the European Commission (EC) allocates relatively little of Europe's science funding.

$308.2 bn

European Commission 3%
Germany 27%
France 16%
UK 13%
Italy 8%
Spain 7%
Sweden 4%
Netherlands 4%
Other 18%

Annual spending 2009. EC funding averaged from FP7 budget.
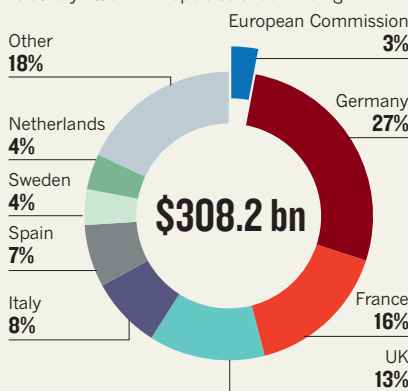
enough to respond quickly to emerging issues", features that Makarow feels neither the ESF nor EUROHORCs has. Being located outside Brussels and holding infrequent meetings have proved disadvantageous to both organizations. And, unlike ScienceEurope, the ESF had no mandate to speak on behalf of its members.

ScienceEurope's pilot board will flesh out the organization's structure and strategy over the coming months. Boyle says that they plan to set up committees covering all research disciplines to guide the organization's activities. They also hope to hold large annual meetings to bring together the plethora of other European and international science and university groups to discuss strategies and debate priorities.

One of ScienceEurope's key goals is to help to build the European Research Area (ERA), a long-cherished ideal within the European Union that would allow researchers to move

freely across borders, taking their funding with them. ScienceEurope's membership base gives it strong links with research policy-makers at national levels, which could help it tackle the thornier problems of the ERA, such as transferring scientists' pensions from one country to another.

Makarow also hopes that ScienceEurope can shift the balance of power in science policy-making away from the European Commission and back towards scientists in member states. She points out that the European Commission manages a tiny fraction of the funds spent annually across Europe (see graph), yet leads the debate on the direction of science policy. "The balance is not right," she says.

Ernst Rietschel, former president of the Leibniz Association of German research institutes, agrees that European scientists need better-coordinated representation. But he and others are concerned that ScienceEurope will not be influential enough because, unlike the ESF, it will not disburse research funding. To have clout you "need money", he observes.

Jean-Pierre Henriet, a geologist and emeritus professor at the University of Ghent in Belgium, is angry that the merger will result in the ESF ending its funding of collaborative research projects. Its annual budget previously provided more than €100 million (US$144 million), collected from member states.

The ESF grants provide "essential funding" for young scientists, who develop contacts and learn networking and leadership skills in collaborative projects, he says, and losing the grants will leave a "major gap" in the research-funding landscape.
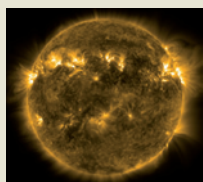
"I understand their concerns," says Makarow, "and hope that other instruments that provide funding at a European level will be developed to fill the gap". ∎

MORE ONLINE

TOP STORY
Model explains chain of linked solar eruptions
go.nature.com/loth7k

OTHER NEWS
● Britain to launch project to provide personalized medicine go.nature.com/7gwyqu
● Vaccines given a clean bill of health go.nature.com/iw5ijz
● Asteroid visit finds familiar dust go.nature.com/z6vlbu

ON THE BLOG
Report on alleged scientific misconduct at Oak Ridge National Laboratory to remain secret, court rules go.nature.com/9onsa9

INVESTIGATION CLOSED

SDO

WORKING WEEKENDS.
LEAVING AT MIDNIGHT.
FRIDAY EVENING MEETINGS.
DOES SCIENCE COME OUT THE WINNER?

THE 24/7 LAB

BY HEIDI LEDFORD

It's just about midnight on a hot Friday night in July, Enrique Iglesias' '*Dirty Dancer*' is on the radio, and 26-year-old graduate student Sagar Shah is starting to look winded. The problem, he says, is not how late it is, or even that he has spent the past three hours working in a cramped sterile cell-culture hood. The problem is that the routine cell-culture maintenance he is doing, bathing his collection of rare human tumour cells with fresh medium, produces no data. And a lack of data, says Sagar, makes him "hungry" for it.

Next to Sagar, Lyonell Kone, a 22-year-old student, rises from another sterile hood and heads for the microscope, jostling his lab-mate Nathaniel Tippens out of the way. He squints at his cultures, checking to make sure the cells are growing at the right density. Satisfied, he backs away, gingerly places his flasks in an incubator, rubs his eyes and stretches. He's finished for the night.

The weary waltz within this cramped cell-culture room is the only flicker of activity at this hour in the Koch Cancer Research Building at Johns Hopkins University in Baltimore, Maryland. It's the Friday before the 4 July holiday, and even the night cleaners quit hours ago, leaving behind the faint smell of disinfectant and the occasional haunting beep of an autoclave echoing down silent hallways. But these members of neurosurgeon Alfredo Quiñones-Hinojosa's laboratory are accustomed to being the last out of the building. In a lab where the boss calls you at 6 a.m., schedules Friday evening lab meetings that can stretch past 10 p.m., and routinely expects you to work over Christmas, sticking it out until midnight on a holiday weekend is nothing unusual.

Many labs are renowned for their intense work ethic and long hours. When I set out to profile such a laboratory, I wanted to find out who is drawn to these environments, what it is really like to work there and whether long hours lead to more or better science. I approached eleven laboratories with reputations for being extremely hard-working. Ten principal investigators turned me down, some expressing a fear of being seen as 'slave-drivers'.

Number eleven — Quiñones-Hinojosa — had no such qualms. His work ethic is no secret: a 2007 essay in the *New England Journal of Medicine*[1] and several television and newspaper reports have traced his path from 19-year-old illegal immigrant from Mexico, labouring in the fields of California, to neurosurgeon at one of the United States' leading research hospitals. He did not get there by working 9 to 5.

Quiñones-Hinojosa fondly recalls the long nights he worked alone in the laboratory as an undergraduate at the University of California, Berkeley, and again as a medical student at Harvard University in Cambridge, Massachusetts. When he was a resident at the University of California, San Francisco, his three young children thought he lived in the hospital. In effect he did, putting in 140 hours a week and grabbing 10-minute naps when he could. Quiñones-Hinojosa credits his professional rise to his resilience and a seemingly limitless capacity for hard work. "When you go that extra step, you are training your brain like an athlete," he says. And the fact that his group has published 113 articles in the past six years and holds 13 funding grants is not, he says, because he is brighter or better connected than colleagues. "It's just a matter of volume," he says. "The key is we submit a couple of dozen grant applications a year, and we learn from our mistakes."

And so, at ease with hard work and the media and steeped in the long-hours culture of medicine, Quiñones-Hinojosa eagerly welcomed me into his research laboratory. "I would be delighted," he said.

The morning I arrived — at 8 a.m. sharp — Quiñones-Hinojosa insisted that I observe his first surgery of the day. He and his resident,

Alfredo Quiñones-Hinojosa (centre, with surgery team) selects lab members who expect long hours and motivates them by inviting patients to lab meetings.

Shaan Raza, were removing a pituitary tumour from a 54-year-old woman. The operating room is an extension of his laboratory, Quiñones-Hinojosa explained: it is there that he collects the tissue samples that his staff — with the patients' consent — will immortalize in cell culture. They are the grist for the lab's studies of how cancer stem cells fuel brain-tumour development and how tumour cells spread through the brain.

Quiñones-Hinojosa had gone to bed at 1 a.m. the night before, and was up again at 5. Walking to surgery, he passes by Kone, who is climbing the stairs to the lab. "You ready to rock 'n' roll?" Quiñones-Hinojosa asks reflexively as he walks by. Then he glances at the time and a mischievous smile darts across his face. "Hey, it's 10 a.m.," he calls over his shoulder, never breaking stride. "What are you doing coming in at 10 a.m.?"

As we walk out of the building Quiñones-Hinojosa nudges me with his elbow and laughs: "See, now he's going to go back to the lab and tell everyone, 'Dr. Q caught me coming in at 10 a.m.!'" (Lab members, who in fact mostly arrive after 9 a.m., confirmed that Kone did exactly that.)

Quiñones-Hinojosa is gregarious and charming, with an infectious energy and a habit of advertising his humility. But he also knows how intimidating he can be to the people who work for him, and he's not afraid to capitalize on that. In 2007, just two years after he started at Hopkins, he rounded a corner in the cafeteria and saw his lab members sitting at a table, talking and laughing. When they caught sight of him, he says, they stopped, stood up, and went straight back to the lab.

Quiñones-Hinojosa has another way to keep his lab motivated. Every so often, he asks a cancer patient or his or her family to join the lab meeting. It is a chance for the patients to learn about the research being done with their tumours. And for the lab, it is a reminder of the urgency of their work. Quiñones-Hinojosa draws out each patient's personal story: how they found out they had cancer, how they felt when they got the news and how it has impacted on their family. Being confronted with all this can be a shock for researchers without medical training. "You can see it in their faces," says Hugo Guerrero-Cázares, a research associate in the lab. "When someone says 'I'm going to die in six months', it really hits them."

Back in the operating room, nurses and surgeons buzz about setting up equipment around the unconscious patient. Pituitary tumours can nestle between the two carotid arteries that supply the brain with blood, making the growths exquisitely difficult to remove. (Quiñones-Hinojosa says he woke up last night worrying about the operation and spent two hours practising every move of the surgery in his mind before nodding off again.) Normally the tumours are about the size of a pea; this one is closer to a golf ball. Quiñones-Hinojosa and Raza meticulously scoop out the tumour piece by piece.

The surgery seems to be a success. Quiñones-Hinojosa steps back from the patient and makes sure the sample is labelled and stored appropriately on ice. He checks with the pathologists down the hall to make sure it includes the tumour tissue he wants, then sends it on to his lab: sample 872 in his collection.

FAST FOOD

In the laboratory, near lunchtime, endocrinology research fellow Nestoras Mathioudakis prepares the tissue in a sterile tissue-culture hood. While the cells are incubating with an enzyme that destroys contaminating red blood cells, he dashes out to eat a frozen meal. He practically lives on them, he says, but worries that the high salt content may be giving him searing headaches. One day, after eating about five

frozen dinners, he sat down at a microscope and found it difficult to focus his right eye. Still, the meals are cheap, fast and a way to grab food without leaving the lab, and Mathioudakis predicts that today will be a busy, multi-frozen-dinner day.

He doesn't really mind. "Only people with a certain type of personality would stay in a lab like this," says Guerrero-Cázares, who has worked there for four years. The night before I arrived, Quiñones-Hinojosa was checking his e-mail on his way home when he noticed a message from a medical student at Rosalind Franklin University of Medicine and Science in Chicago, Illinois, who wanted to work in the laboratory. Quiñones-Hinojosa receives several such enquiries a day, but something about this student — Joshua Bakhsheshian — caught his eye. He fired back a message: give me a number at which I can reach you at 6 a.m.. It was midnight. A minute later he had his reply.

At 6:02 a.m. Quiñones-Hinojosa called Bakhsheshian. "I laid it on so thick for this guy," Quiñones-Hinojosa crooned later that morning. "I said, 'You've seen me on TV, you think I'm so nice. But you come into my lab, you're going to work. The people in my lab, they work 24 hours a day. They're here over Christmas and New Year writing grants, and you will be, too.'"

"That's music to my ears," replied Bakhsheshian, who later told me he had never expected such a speedy reply from the busy surgeon, and had stayed up much of the night frantically studying the lab's publications. (Quiñones-Hinojosa later offered him a spot in the lab if Bakhsheshian could get a fellowship.)

Not everyone whom Quiñones-Hinojosa selects adapts well to the rigours of his laboratory. Research fellow David Chesler, a neurosurgery resident at the University of Maryland in College Park with a PhD in neuroimmunology and circles under his eyes, recalls a technician who "wasn't keeping up" — and Guerrero-Cázares recounts the tale of a colleague who simply stopped coming to the Friday night lab meetings. Both left the lab. Quiñones-Hinojosa says that he asked them to leave "very nicely", and helped them to find positions elsewhere.

Still, Quiñones-Hinojosa's technique of screening for work habits and personality traits may be one reason why the lab runs so smoothly, despite its intensity. Pierre Azoulay, associate professor of strategy at the Massachusetts Institute of Technology's Sloan School of Management in Cambridge, says that asking an employee to work long hours can backfire if that person is used to operating differently. "Unless you select your trainees very carefully on those criteria — which I wager most principal investigators don't — there would presumably be deleterious effects."

Another key is autonomy. Many members of the Quiñones-Hinojosa lab develop their own projects, and write the grant applications to fund them. They express a proud sense of ownership when it comes to their work. And despite the 6 a.m. phone calls from the boss — made during his commute to the hospital — they say they feel reasonably free to set their own schedules. Shah says that 20-hour days are not uncommon for him. But "I don't believe in clocking in and clocking out," he says. "I could do that at Walmart and get overtime."

That freedom is essential to keeping researchers happy and productive, says Azoulay. "Science is a harsh mistress," he says. "I think relatively few scientists are expecting 9-to-5 jobs. But they are expecting autonomy, and a principal investigator that violates that expectation could potentially run into problems."

So far, Quiñones-Hinojosa's lab seems relatively problem-free. But are the long hours and personal sacrifices worth it, for the lab members and for science? In 2004, Steven Stack, a sociologist at Wayne State University in Detroit, Michigan, published an analysis of survey data collected by the US National Research Council on 11,231 PhD scientists and engineers working in academia[2]. He found that the average scientist worked about 50 hours a week, and in general the more hours an individual put in, the more publications he or she cranked out.

Quiñones-Hinojosa's lab seems to fit that mould. Of the 113 articles he has published since he launched the lab in 2005, most are from a small 'dry' laboratory working on clinical outcomes in cancer. His 27-person 'wet' lab has published 29. Overall, his h index — a measure of productivity that factors in the number of articles published and how often they are cited — is 27, compared with 10.7 for US neurosurgeons at the same associate professor level[3]. Quiñones-Hinojosa also notes that it takes researchers in his department an average of 15 years to be promoted to full professor. He was recommended for a full professorship this year, after just six.

Biochemist Philip Cohen of the University of Dundee, UK, says that of the 70 postdocs and nearly 50 students he has supervised during his career, the most successful were those who put in long hours and worked efficiently. Cohen frets that the lab culture is changing. "Everyone's told not to stress themselves or overdo things, and I could not disagree more," he says. "I'm afraid they're losing all the fun in life if they don't really push themselves to the limit."

But not everyone agrees that more hours yield more results. Dean Simonton, a psychology researcher at the University of California, Davis, who has studied scientific creativity, says that the pressure for publications, grants and tenure may have created a single-minded, "monastic" culture in science. But some research suggests that highly creative scientists tend to have broader interests and more hobbies than their less creative colleagues, he says. Chemist Stephen Buchwald of the Massachusetts Institute of Technology urges the members of his lab to take a month's holiday every year, and not to think about work when they're gone. "The fact is, I want people to be able to think," he says. "If they're completely beaten down, they're not going to be very creative." His approach does not seem to have hurt productivity: Thomson Reuters declared Buchwald one of the most highly cited chemists from 1999 to 2009, with an average of more than 86 citations for his 171 papers.

An intense work schedule also comes with personal costs that can be hard to measure. "The area in which I have failed the most is as a father," Quiñones-Hinojosa readily admits. It is something he is trying to correct, by spending more time with his kids and shuttling them to swimming lessons (although phoning lab members on the way).

And postdoc Pragathi Achanta looks wistful when she talks about her niece in India, who was six months old the last time Achanta saw her — now she's nearly five. Achanta has been working on grant applications over the holidays, and hasn't had time to visit her family.

Now, at 8 p.m. on Friday 1 July, Achanta is taking advantage of the unusually short lab meeting to prepare surgical tools for a mouse experiment to model the effects of radiotherapy on stem cells. She wants to be ready so that she can complete the surgeries quickly on Saturday morning before she leaves to help teach a course at Cold Spring Harbor Laboratory in New York. Later this year, grant schedule allowing, she hopes to travel to India to see her niece at last. But she admits to being nervous about broaching the subject with the boss.

Quiñones-Hinojosa, though, says that he has nothing against holidays. "Vacations are great," he says. "Take a weekend off." ∎

"THE PEOPLE IN MY LAB, THEY WORK 24 HOURS A DAY. THEY'RE HERE OVER CHRISTMAS AND NEW YEAR WRITING GRANT APPLICATIONS."

Heidi Ledford is a reporter for Nature in Cambridge, Massachusetts.

1. Quiñones-Hinojosa, A. N. Engl. J. Med. 357, 529–531 (2007).
2. Stack, S. Res. Higher Ed. 45, 891–920 (2004).
3. Lee, J., Kraus, K. L. & Couldwell, W. T. J. Neurosurg. 111, 387–392 (2009).

TAKING AIM AT FREE WILL

Scientists think they can prove that free will is an illusion. Philosophers are urging them to think again.

BY KERRI SMITH

FLATLINER/GETTY

THE EXPERIMENT HELPED TO CHANGE JOHN-DYLAN HAYNES'S OUTLOOK ON LIFE.
In 2007, Haynes, a neuroscientist at the Bernstein Center for Computational Neuroscience in Berlin, put people into a brain scanner in which a display screen flashed a succession of random letters[1]. He told them to press a button with either their right or left index fingers whenever they felt the urge, and to remember the letter that was showing on the screen when they made the decision. The experiment used functional magnetic resonance imaging (fMRI) to reveal brain activity in real time as the volunteers chose to use their right or left hands. The results were quite a surprise.

▶

▶ "The first thought we had was 'we have to check if this is real,'" says Haynes. "We came up with more sanity checks than I've ever seen in any other study before."

The conscious decision to push the button was made about a second before the actual act, but the team discovered that a pattern of brain activity seemed to predict that decision by as many as seven seconds. Long before the subjects were even aware of making a choice, it seems, their brains had already decided.

As humans, we like to think that our decisions are under our conscious control — that we have free will. Philosophers have debated that concept for centuries, and now Haynes and other experimental neuroscientists are raising a new challenge. They argue that consciousness of a decision may be a mere biochemical afterthought, with no influence whatsoever on a person's actions. According to this logic, they say, free will is an illusion. "We feel we choose, but we don't," says Patrick Haggard, a neuroscientist at University College London.

You may have thought you decided whether to have tea or coffee this morning, for example, but the decision may have been made long before you were aware of it. For Haynes, this is unsettling. "I'll be very honest, I find it very difficult to deal with this," he says. "How can I call a will 'mine' if I don't even know when it occurred and what it has decided to do?"

THOUGHT EXPERIMENTS

Philosophers aren't convinced that brain scans can demolish free will so easily. Some have questioned the neuroscientists' results and interpretations, arguing that the researchers have not quite grasped the concept that they say they are debunking. Many more don't engage with scientists at all. "Neuroscientists and philosophers talk past each other," says Walter Glannon, a philosopher at the University of Calgary in Canada, who has interests in neuroscience, ethics and free will.

There are some signs that this is beginning to change. This month, a raft of projects will get under way as part of Big Questions in Free Will, a four-year, US$4.4-million programme funded by the John Templeton Foundation in West Conshohocken, Pennsylvania, which supports research bridging theology, philosophy and natural science. Some say that, with refined experiments, neuroscience could help researchers to identify the physical processes underlying conscious intention and to better understand the brain activity that precedes it. And if unconscious brain activity could be found to predict decisions perfectly, the work really could rattle the notion of free will. "It's possible that what are now correlations could at some point become causal connections between brain mechanisms and behaviours," says Glannon. "If that were the case, then it would threaten free will, on any definition by any philosopher."

Haynes wasn't the first neuroscientist to explore unconscious decision-making. In the 1980s, Benjamin Libet, a neuropsychologist at the University of California, San Francisco, rigged up study participants to an electro-encephalogram (EEG) and asked them to watch a clock face with a dot sweeping around it[2]. When the participants felt the urge to move a finger, they had to note the dot's position. Libet recorded brain activity several hundred milliseconds before people expressed their conscious intention to move.

Libet's result was controversial. Critics said that the clock was distracting, and the report of a conscious decision was too subjective. Neuroscience experiments usually have controllable inputs — show someone a picture at a precise moment, and then look for reactions in the brain. When the input is the participant's conscious intention to move, however, they subjectively decide on its timing. Moreover, critics weren't convinced that the activity seen by Libet before a conscious decision was sufficient to cause the decision — it could just have been the brain gearing up to decide and then move.

Haynes's 2008 study[1] modernized the earlier experiment: where Libet's EEG technique could look at only a limited area of brain activity, Haynes's fMRI set-up could survey the whole brain; and where Libet's participants decided simply on when to move, Haynes's test forced them to decide between two alternatives. But critics still picked holes, pointing out that Haynes and his team could predict a left or right button press with only 60% accuracy at best. Although better than chance, this isn't enough to claim that you can see the brain making its mind up before conscious awareness, argues Adina Roskies, a neuroscientist and philosopher who works on free will at Dartmouth College in Hanover, New Hampshire. Besides, "all it suggests is that there are some physical factors that influence decision-making", which shouldn't be surprising. Philosophers who know about the science, she adds, don't think this sort of study is good evidence for the absence of free will, because the experiments are caricatures of decision-making. Even the seemingly simple decision of whether to have tea or coffee is more complex than deciding whether to push a button with one hand or the other.

Haynes stands by his interpretation, and has replicated and refined his results in two studies.

One uses more accurate scanning techniques[3] to confirm the roles of the brain regions implicated in his previous work. In the other, which is yet to be published, Haynes and his team asked subjects to add or subtract two numbers from a series being presented on a screen. Deciding whether to add or subtract reflects a more complex intention than that of whether to push a button, and Haynes argues that it is a more realistic model for everyday decisions. Even in

> ## HOW CAN I CALL A WILL 'MINE' IF I DON'T EVEN KNOW WHEN IT OCCURRED AND WHAT IT HAS DECIDED TO DO?

this more abstract task, the researchers detected activity up to four seconds before the subjects were conscious of deciding, Haynes says.

Some researchers have literally gone deeper into the brain. One of those is Itzhak Fried, a neuroscientist and surgeon at the University of California, Los Angeles, and the Tel Aviv Medical Center in Israel. He studied individuals with electrodes implanted in their brains as part of a surgical procedure to treat epilepsy[4]. Recording from single neurons in this way gives scientists a much more precise picture of brain activity than fMRI or EEG. Fried's experiments showed that there was activity in individual neurons of particular brain areas about a second and a half before the subject made a conscious decision to press a button. With about 700 milliseconds to go, the researchers could predict the timing of that decision with more than 80% accuracy. "At some point, things that are predetermined are admitted into consciousness," says Fried. The conscious will might be added on to a decision at a later stage, he suggests.

MATERIAL GAINS

Philosophers question the assumptions underlying such interpretations. "Part of what's driving some of these conclusions is the thought that free will has to be spiritual or involve souls or something," says Al Mele, a philosopher at Florida State University in Tallahassee. If neuroscientists find unconscious neural activity that drives decision-making, the troublesome concept of mind as separate from body disappears, as does free will. This 'dualist' conception of free will is an easy target for neuroscientists to knock down, says Glannon. "Neatly dividing mind and brain makes it easier for neuroscientists to drive a wedge between them," he adds.

The trouble is, most current philosophers don't think about free

⊃ NATURE.COM
To listen to a podcast about neuroscience and free will, visit:
go.nature.com/ihlh5z

will like that, says Mele. Many are materialists — believing that everything has a physical basis, and decisions and actions come from brain activity. So scientists are weighing in on a notion that philosophers consider irrelevant.

Nowadays, says Mele, the majority of philosophers are comfortable with the idea that people can make rational decisions in a deterministic universe. They debate the interplay between freedom and determinism — the

says. Some informal meetings have already begun. Roskies, who is funded through the programme, plans to spend time this year in the lab of Michael Shadlen, a neurophysiologist at the University of Washington in Seattle who works on decision-making in the primate brain. "We're going to hammer on each other until we really understand the other person's point of view, and convince one or other of us that we're wrong," she says.

Foundation in Chicago, Illinois, suggests that the research could help to identify an individual's level of responsibility. "What we are interested in is how neuroscience can give us a more granulated view of how people vary in their ability to control their behaviour," says Jones. That could affect the severity of a sentence, for example.

The answers could also end up influencing people's behaviour. In 2008, Kathleen Vohs, a social psychologist at the University of Minnesota in Minneapolis, and her colleague Jonathan Schooler, a psychologist now at the University of California, Santa Barbara, published a study[5] on how people behave when they are prompted to think that determinism is true. They asked their subjects to read one of two passages: one suggesting that behaviour boils down to environmental or genetic factors not under personal control; the other neutral about what influences behaviour. The participants then did a few maths problems on a computer. But just before the test started, they were informed that because of a glitch in the computer it occasionally displayed the answer by accident; if this happened, they were to click it away without looking. Those who had read the deterministic message were more likely to cheat on the test. "Perhaps, denying free will simply provides the ultimate excuse to behave as one likes," Vohs and Schooler suggested.

Haynes's research and its possible implications have certainly had an effect on how he thinks. He remembers being on a plane on his way to a conference and having an epiphany. "Suddenly I had this big vision about the whole deterministic universe, myself, my place in it and all these different points where we believe we're making decisions just reflecting some causal flow." But he couldn't maintain this image of a world without free will for long. "As soon as you start interpreting people's behaviours in your day-to-day life, it's virtually impossible to keep hold of," he says.

Fried, too, finds it impossible to keep determinism at the top of his mind. "I don't think about it every day. I certainly don't think about it when I operate on the human brain."

Mele is hopeful that other philosophers will become better acquainted with the science of conscious intention. And where philosophy is concerned, he says, scientists would do well to soften their stance. "It's not as though the task of neuroscientists who work on free will has to be to show there isn't any." ∎

THAT THERE ARE SOME PHYSICAL FACTORS THAT INFLUENCE DECISION-MAKING SHOULDN'T BE A SURPRISE.

theory that everything is predestined, either by fate or by physical laws — but Roskies says that results from neuroscience can't yet settle that debate. They may speak to the predictability of actions, but not to the issue of determinism.

Neuroscientists also sometimes have misconceptions about their own field, says Michael Gazzaniga, a neuroscientist at the University of California, Santa Barbara. In particular, scientists tend to see preparatory brain activity as proceeding stepwise, one bit at a time, to a final decision. He suggests that researchers should instead think of processes working in parallel, in a complex network with interactions happening continually. The time at which one becomes aware of a decision is thus not as important as some have thought.

BATTLE OF WILLS
There are conceptual issues — and then there is semantics. "What would really help is if scientists and philosophers could come to an agreement on what free will means," says Glannon. Even within philosophy, definitions of free will don't always match up. Some philosophers define it as the ability to make rational decisions in the absence of coercion. Some definitions place it in cosmic context: at the moment of decision, given everything that's happened in the past, it is possible to reach a different decision. Others stick to the idea that a non-physical 'soul' is directing decisions.

Neuroscience could contribute directly to tidying up definitions, or adding an empirical dimension to them. It might lead to a deeper, better understanding of what freely willing something involves, or refine views of what conscious intention is, says Roskies.

Mele is directing the Templeton Foundation project that is beginning to bring philosophers and neuroscientists together. "I think if we do a new generation of studies with better design, we'll get better evidence about what goes on in the brain when people make decisions," he

Haggard has Templeton funding for a project in which he aims to provide a way to objectively determine the timing of conscious decisions and actions, rather than rely on subjective reports. His team plans to devise an experimental set-up in which people play a competitive game against a computer while their brain activity is decoded.

Another project, run by Christof Koch, a bioengineer at the California Institute of Technology in Pasadena, will use techniques similar to Fried's to examine the responses of individual neurons when people use reason to make decisions. His team hopes to measure how much weight people give to different bits of information when they decide.

Philosophers are willing to admit that neuroscience could one day trouble the concept of free will. Imagine a situation (philosophers like to do this) in which researchers could always predict what someone would decide from their brain activity, before the subject became aware of their decision. "If that turned out to be true, that would be a threat to free will," says Mele. Still, even those who have perhaps prematurely proclaimed the death of free will agree that such results would have to be replicated on many different levels of decision-making. Pressing a button or playing a game is far removed from making a cup of tea, running for president or committing a crime.

The practical effects of demolishing free will are hard to predict. Biological determinism doesn't hold up as a defence in law. Legal scholars aren't ready to ditch the principle of personal responsibility. "The law has to be based on the idea that people are responsible for their actions, except in exceptional circumstances," says Nicholas Mackintosh, director of a project on neuroscience and the law run by the Royal Society in London.

Owen Jones, a law professor at Vanderbilt University in Nashville, Tennessee, who directs a similar project funded by the MacArthur

Kerri Smith *is editor of the* Nature Podcast, *and is based in London.*

1. Soon, C. S., Brass, M., Heinze, H.-J. & Haynes, J.-D. *Nature Neurosci.* **11,** 543–545 (2008).
2. Libet, B., Gleason, C. A., Wright, E. W. & Pearl, D. K. *Brain* **106,** 623–642 (1983).
3. Bode, S. *et al. PLoS ONE* **6,** e21612 (2011).
4. Fried, I., Mukamel, R. & Kreiman, G. *Neuron* **69,** 548–562 (2011).
5. Vohs, K. D. & Schooler, J. W. *Psychol. Sci.* **19,** 49–54 (2008).

COMMENT

ILLUSTRATION BY MICHAEL KIRKHAM

A healthy work–life balance can enhance research

Scientists should make time for play to complement their intense work, maintain creativity and keep the ideas flowing, argues **Julie Overbaugh**.

Striking a balance between work, family, friends and leisure is often hard in science. But there must be room for those who want this balance, otherwise creative people with the potential to make significant contributions to scientific discovery will be excluded.

In my college years at the University of Connecticut in Storrs, I juggled a chemistry major with playing on First Division college basketball and tennis teams. And, ingrained from my Irish upbringing, I felt that college was also meant to be a time where some evenings were spent in the pub with friends. I brought art classes into the mix in graduate school. From these experiences, I learned to value efficiency, balance and teamwork, all of which have influenced my approach to running my research programme. Today I direct a US lab of about 15 people and share responsibility for a much larger international team focused on HIV prevention research.

During my training, I experienced many different lab styles and, in the process, realized that I was not cut out for clocking long hours for the rest of my life. I decided to try to do science in a manner that I could integrate into my life, leaving time for an extensive network of friends and family and other interests, and accept that there would be other paths if this one did not work — a crucial mindset for not succumbing to a frenzied work life.

Twenty years on, I have found that spending long hours in the lab or at the computer does not necessarily promote the creative thinking that is integral to scientific discovery. In fact, I have many of my best ideas while walking the dogs in the morning, riding my bike home from work or weekending in the mountains.

QUALITY NOT QUANTITY

I have no objection to people who choose to do long hours in the lab, but I have also never expected lab members to give up other ▶

▶ aspects of their lives to pursue science. One of my team recently noted: "You expect high-quality work — but you do not measure someone's efforts by time spent in the lab in the evenings or on weekends, rather you measure them by the quality of their data."

To be a successful scientist there are times when it is important to pull out all the stops — when a big grant deadline is looming or a high-impact paper is wrapping up. Sometimes, when we are competing with other labs on an exciting story, I briefly imagine locking everyone in the lab to try to push for results more quickly.

By resisting the temptation to drive my group so hard, we might have ended up with the second or third paper on a topic in a less prestigious journal on occasion. But our contributions are clearly recognized. For example: I have given plenary talks at every major HIV meeting; we have received two National Institutes of Health (NIH) MERIT awards; former trainees are faculty members at Harvard University in Cambridge, Massachusetts, Stanford University in California, Baylor College of Medicine in Houston, Texas, and many other top research institutions; others are successful government scientists at the NIH and the Centers for Disease Control and Prevention.

More importantly, in many cases, we solved a problem more effectively, and thus gained some advantage, because people in the lab were less stressed by long hours, constant demands and excessive expectations. In my view, an unremitting pace with no time to step back leads, over the long term, to a fatigued and unhappy team that is not operating at its best.

TIME FOR TEA

For the past two decades, I have worked in a highly productive interdisciplinary and international collaboration with scientists from many cultures. We are committed to reducing the burden of HIV, particularly in developing countries, and many on the team regularly see the devastation of HIV in clinical practice.

But colleagues also have a range of demands on their time, including young faculty members who juggle research with clinical work and raising children or helping ageing parents. The original collaboration was driven by the needs of the science; it is sustained by *esprit de corps*. The glue is our appreciation of the specific needs and lifestyle choices of each group member, including our trainees.

I am also fortunate to work at an institution where the focus is on our contributions, not our hours. When a colleague recently suffered a devastating accident, many faculty members and staff were regularly by her hospital bed for months, placing their presence there at higher priority than their work. Thus, it is noteworthy that, among 209 faculty members, we have three Nobel laureates and numerous others who have received prestigious awards (go.nature.com/zebxjt).

Balancing work and other aspects of life is becoming harder in science, as in other professions. We are expected to be constantly responsive to e-mail, alert to rapid online publications, to manage increasing administrative and regulatory demands, and to devote more and more time to securing funding.

Tighter funding and ease of travel have made 'being on tour' an integral part of being the lab chief or principal investigator (PI), often to the exclusion of spending time with the team. Being a lab head can be more PR than PI, with the pressure to 'sell your goods' on the road. I realized that attending just a few meetings a year was enough when I reached the point that I had heard my colleagues speak so many times that I could have given some of their talks myself, and they mine.

Meetings are wonderful for networking and exchanging new findings and ideas, but they can have diminishing returns in terms of promoting creativity or seeding a breakthrough. We should re-evaluate the expanding repertoire of conferences in some fields to determine if the scientific exchange genuinely offsets the demands of constant travel.

More generally, time to think deeply about scientific problems is becoming increasingly rare. Indeed, with the rapid advance of new technologies, many studies simply apply new methods to old problems, often getting a more refined, but similar answer. Although it is comforting to see the consistency in conclusions, the true advances in such cases may be small.

About 25 years ago, I was the lone trainee in a lab of a prominent senior scientist. We spent many afternoons drinking tea and discussing diverse scientific issues. Current scientific discovery would benefit from reinstating this mostly bygone tradition. Imagine, further, if labs everywhere shut down all experiments for a week each year to have everyone read recent — and old — literature and discuss ideas over a favourite beverage. I predict that new ideas would emerge and people would be energized.

It is a privilege to have work that is challenging and satisfying, and that contributes to society and helps advance knowledge and training. For some, a total focus on research and the thrill of discovery is fulfilling. For others, originality, productivity and overall happiness in life demands a different mix. There must be room for both types of people in science, as each has the potential to make important contributions. ■ **SEE EDITORIAL P.5 AND NEWS FEATURE P.20**

Julie Overbaugh *is in the Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. e-mail: joverbau@fhcrc.org*
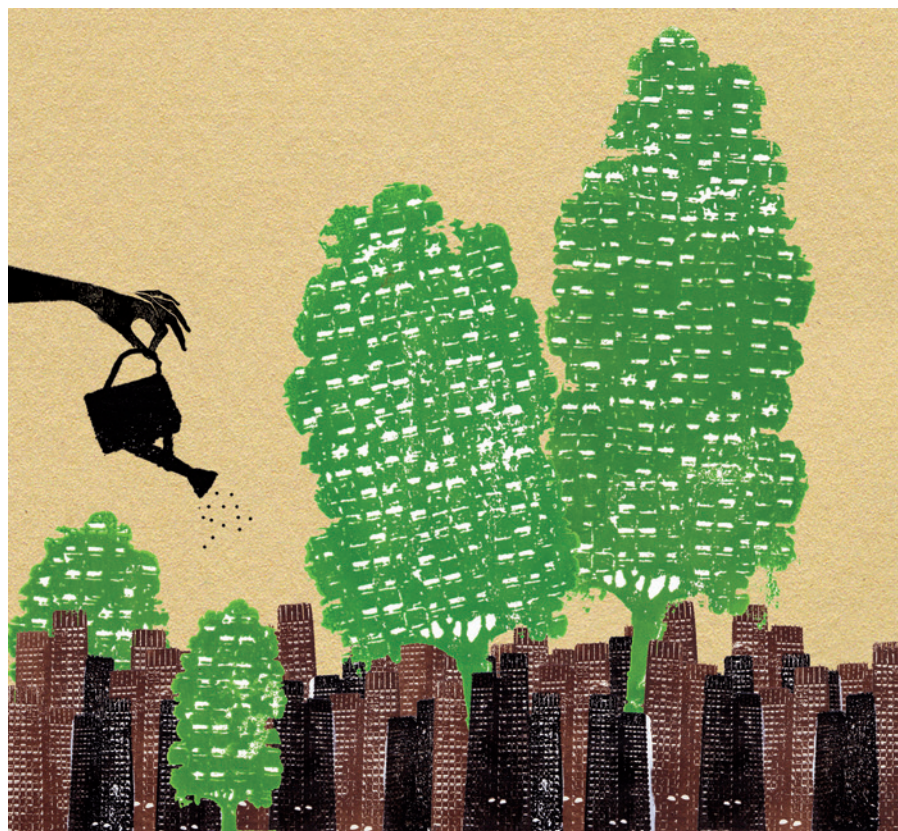
ECOLOGY

Redefining nature

Shahid Naeem compares two books that call for us to embrace the influence of humans on ecosystems.

In Mary Shelley's novel, Victor Frankenstein assembled parts of dead humans into a creature he brought to life by unorthodox scientific methods. The resulting being, although human in form and function, was seen as a monster. If we were to assemble a thing made up of plants, animals and microbes, and then breathe ecological life into it, would it be nature or an unnatural monster? Two authors examine this issue: Emma Marris in *Rambunctious Garden* and Nigel Dudley in *Authenticity in Nature*.

If one defines 'natural' nature as assemblages of native species that occupy large, unbroken areas without obvious human influence, and that exhibit normal (1,000-year average) rates of extinction, origination and ecosystem function, then there is no such thing on Earth today. The destruction of the natural world began some 2 million years ago, with burning and hunting by *Homo erectus*. In the past 50,000 years, *Homo sapiens* took these skills to new heights, culminating in the Industrial Revolution, in which some

---

Rambunctious Garden: Saving Nature in a Post-wild World
EMMA MARRIS
Bloomsbury: 2011. 288 pp. £20, $25

Authenticity in Nature: Making Choices about the Naturalness of Ecosystems
NIGEL DUDLEY
Earthscan: 2011. 256 pp. £19.99, $34.95

---

25 million square kilometres of grasslands were burned, 12 million square kilometres of forests vanished, and floral and faunal extinctions skyrocketed worldwide. If the loss of the natural world is nothing new, does it matter? Both authors think not.

Dudley opens *Authenticity in Nature* by recounting his boyhood experiences at the Attenborough Nature Reserve near Nottingham, UK. Now leased by a mining company to a wildlife trust, it was for centuries grazing land. To visitors, it is nature. Marris goes a step further in

Rambunctious Garden, suggesting that birds and bees on New York's Fifth Avenue constitute nature. I could go one better: my formative childhood experiences of nature were the dioramas of the American Museum of Natural History in New York. My mind's eye breathed life into the stuffed, encased creatures; for me, they defined nature.

Both authors review what is natural and unnatural from an ecological perspective, but their approaches are different. Marris, an accomplished science writer who regularly writes for *Nature*, follows the tactic of David Quammen (*The Song of the Dodo*; 1996), Stuart Pimm (*The World According to Pimm*; 2001) and Jonathan Weiner (*The Beak of the Finch*; 1994). In this, her first book, she brings together her many travels and encounters to create a Kerouac-like journey through which her thesis emerges. She posits that significant parts of the modern world are or will become "rambunctious gardens" — unruly entanglements of weedy species that follow in our wake. Because nature is ever changing, on the scale of Earth's history, rambunctious gardens are as legitimate as any other manifestation of nature. We should embrace our creations, not shun them as monsters.

Marris makes effective use of the popular voice, but sometimes goes overboard. Phrases such as "North American mammoths went kaput" do injustice to the gravity of such issues. Similarly, frequent references to "ecologists", as in her claim that most ecologists dislike anything that "reeks of mankind", belies the broad spectrum of views that such researchers hold.

Dudley's book is more sombre, similar to Simon Levin's *Fragile Dominion* (1999), E. O. Wilson's *The Future of Life* (2002) or Carolyn Merchant's *Reinventing Eden* (2003). An established environmental scientist, Dudley also uses his travels to bring colour to his writing, but his hypothesis comes from a thoughtful examination of various attempts to define 'natural' and 'wild' — by scientists, philosophers, managers, non-governmental organizations and policy-makers. Dudley demonstrates that there is neither coherence nor consensus as to what constitutes naturalness or wildness. He suggests that we focus instead on "authenticity".

For nature to be authentic, Dudley posits, it need only contain a web of interacting species that provides stable ecosystem functions and services. The islands of Hawaii, for example, are inhabited by humans, have suffered innumerable extinctions and are riddled with non-native species. Yet they remain complex and productive ecosystems — not natural, ▶

> "There is neither coherence nor consensus as to what constitutes naturalness or wildness."

▶ perhaps, but authentic Hawaiian ones. Being authentic does not involve having native or endemic species or being devoid of people, and is thus a more tractable environmental goal than achieving naturalness or wildness.

Neither volume tries to dismiss 'natural nature' as the cause célèbre of conservation; rather, both encourage adding unnatural nature to that which we seek to preserve. Well protected places that are rich in endemics are important, but rambunctious gardens and authentic ecosystems are crucial too.

Take a modern New England forest. Its ecology consists of garlic mustard from Europe, thorny Japanese barberry, introduced earthworms, soil that is enriched with industrial nitrogen, exotic insect pests such as hemlock woolly adelgids, Asian longhorned beetles and emerald ash borers, and an ungodly number of ravenous deer who know no serious predation. It is a rambunctious garden. It is polluted, stunted, diseased, unstable and has no top predators. But it is teeming with life. It is home to surviving native species; exhibits ecosystem functions such as storing carbon and cycling nutrients; and even provides some ecosystem services by stabilizing the hilly slopes and supplying deer for hunters in the region. This is authentic nature in the Anthropocene epoch.

Yet ecological theory still holds, even if it is complex and inconvenient. If nature is any set of interacting species, through which energy flows and nutrients cycle, then it calls for saving too many species and setting aside too much land. Yet I would caution against making intelligent-design-like arguments that would dispense with ecology, to replace it with something simpler. Nature without ecology is like biology without evolution; neither is viable, neither makes sense.

Marris and Dudley challenge us to revisit the definition of nature in our increasingly unnatural world. But modern ecosystems that are haphazardly assembled from the remains of human development are unpredictable and fragile. With one billion people hungry, two billion poor and three billion in desperate need of water, the hopes of humanity rest on conserving, restoring and sustainably managing the services that nature provides. The bigger question is whether the unnatural nature that we have wrought, although familiar in form and function, will save us or prove to be monstrous. ∎

Shahid Naeem is professor of ecology at Columbia University, New York, USA.
e-mail: sn2121@columbia.edu

Building the Large Hadron Collider proved tricky, not least because of fears it would create tiny black holes.

PARTICLE PHYSICS

Inside the collider

Joseph Silk enjoys an eloquent take on the Higgs boson, supersymmetry and the world's largest particle smasher.

Mentioning particle physics may silence many dinner parties, but that has not deterred its funders. By the end of 2010, more than €7 billion (US$10 billion) had been ploughed into the current world-leading machine in experimental particle physics — the Large Hadron Collider (LHC) at CERN, Europe's high-energy physics lab near Geneva, Switzerland. So it behooves the researchers involved to communicate the relevance of the LHC's science goals to the public.

Lisa Randall's *Knocking on Heaven's Door* is the latest attempt to do so. Her eloquent book details the trials and tribulations of the LHC, from conception to implementation, and takes us on a grand tour of the underlying science. Randall, a professor of physics at Harvard University in Cambridge, Massachusetts, and a leading contributor to particle-physics theory, borrows her title from Bob Dylan's soundtrack to the 1973 Sam Peckinpah film, *Pat Garrett and Billy the Kid*. The film is a lament on the death of a gunslinger — and the book's title may be a reference to the prediction that turning on the LHC would result in the destruction of Earth. Fortunately, as Randall describes, this did not happen.

That prediction provides a measure of the LHC's reach, and its hold on the public's imagination. Physicists' ultimate dream is to unify the fundamental interactions of physics. This involves combining

⤷ **NATURE.COM**
Lisa Randall's hyperspace opera:
go.nature.com/ydazso

gravity with quantum theory and, in particular, with the forces that hold particles and atoms together in higher dimensions of space and time.

String theorists, whose ranks purportedly include the greatest brains in physics, have predicted that infinitesimal, very short-lived black holes are the unifying factor, the missing glue. Some have predicted that these microscopic objects could be recreated in sufficiently high-energy particle collisions with a powerful particle accelerator such as the LHC. However, most particle physicists doubt that they will actually see such events — Stephen Hawking predicted that, owing to quantum physical effects, microscopic black holes should decay in a fraction of a nanosecond.

But even Hawking might be fallible. Richard Feynman famously said that "nobody understands quantum mechanics". If Hawking was wrong, an escaping black hole might suck up its surroundings: the LHC itself and Geneva (to which humanity could no doubt adapt) and even Earth. Pursuing this logic, a teacher in Hawaii combined

Knocking on Heaven's Door: How Physics and Scientific Thinking Illuminate the Universe and the Modern World
LISA RANDALL
Bodley Head: 2011.
464 pp. £20

▶ perhaps, but authentic Hawaiian ones. Being authentic does not involve having native or endemic species or being devoid of people, and is thus a more tractable environmental goal than achieving naturalness or wildness.

Neither volume tries to dismiss 'natural nature' as the cause célèbre of conservation; rather, both encourage adding unnatural nature to that which we seek to preserve. Well protected places that are rich in endemics are important, but rambunctious gardens and authentic ecosystems are crucial too.

Take a modern New England forest. Its ecology consists of garlic mustard from Europe, thorny Japanese barberry, introduced earthworms, soil that is enriched with industrial nitrogen, exotic insect pests such as hemlock woolly adelgids, Asian longhorned beetles and emerald ash borers, and an ungodly number of ravenous deer who know no serious predation. It is a rambunctious garden. It is polluted, stunted, diseased, unstable and has no top predators. But it is teeming with life. It is home to surviving native species; exhibits ecosystem functions such as storing carbon and cycling nutrients; and even provides some ecosystem services by stabilizing the hilly slopes and supplying deer for hunters in the region. This is authentic nature in the Anthropocene epoch.

Yet ecological theory still holds, even if it is complex and inconvenient. If nature is any set of interacting species, through which energy flows and nutrients cycle, then it calls for saving too many species and setting aside too much land. Yet I would caution against making intelligent-design-like arguments that would dispense with ecology, to replace it with something simpler. Nature without ecology is like biology without evolution; neither is viable, neither makes sense.

Marris and Dudley challenge us to revisit the definition of nature in our increasingly unnatural world. But modern ecosystems that are haphazardly assembled from the remains of human development are unpredictable and fragile. With one billion people hungry, two billion poor and three billion in desperate need of water, the hopes of humanity rest on conserving, restoring and sustainably managing the services that nature provides. The bigger question is whether the unnatural nature that we have wrought, although familiar in form and function, will save us or prove to be monstrous. ∎

Shahid Naeem *is professor of ecology at Columbia University, New York, USA.*
e-mail: sn2121@columbia.edu

Building the Large Hadron Collider proved tricky, not least because of fears it would create tiny black holes.

C. MARCELLONI/CERN

PARTICLE PHYSICS

Inside the collider

Joseph Silk enjoys an eloquent take on the Higgs boson, supersymmetry and the world's largest particle smasher.

Mentioning particle physics may silence many dinner parties, but that has not deterred its funders. By the end of 2010, more than €7 billion (US$10 billion) had been ploughed into the current world-leading machine in experimental particle physics — the Large Hadron Collider (LHC) at CERN, Europe's high-energy physics lab near Geneva, Switzerland. So it behooves the researchers involved to communicate the relevance of the LHC's science goals to the public.

Lisa Randall's *Knocking on Heaven's Door* is the latest attempt to do so. Her eloquent book details the trials and tribulations of the LHC, from conception to implementation, and takes us on a grand tour of the underlying science. Randall, a professor of physics at Harvard University in Cambridge, Massachusetts, and a leading contributor to particle-physics theory, borrows her title from Bob Dylan's soundtrack to the 1973 Sam Peckinpah film, *Pat Garrett and Billy the Kid*. The film is a lament on the death of a gunslinger — and the book's title may be a reference to the prediction that turning on the LHC would result in the destruction of Earth. Fortunately, as Randall describes, this did not happen.

That prediction provides a measure of the LHC's reach, and its hold on the public's imagination. Physicists' ultimate dream is to unify the fundamental interactions of physics. This involves combining

gravity with quantum theory and, in particular, with the forces that hold particles and atoms together in higher dimensions of space and time.

String theorists, whose ranks purportedly include the greatest brains in physics, have predicted that infinitesimal, very short-lived black holes are the unifying factor, the missing glue. Some have predicted that these microscopic objects could be recreated in sufficiently high-energy particle collisions with a powerful particle accelerator such as the LHC. However, most particle physicists doubt that they will actually see such events — Stephen Hawking predicted that, owing to quantum physical effects, microscopic black holes should decay in a fraction of a nanosecond.

But even Hawking might be fallible. Richard Feynman famously said that "nobody understands quantum mechanics". If Hawking was wrong, an escaping black hole might suck up its surroundings: the LHC itself and Geneva (to which humanity could no doubt adapt) and even Earth. Pursuing this logic, a teacher in Hawaii combined

Knocking on Heaven's Door: How Physics and Scientific Thinking Illuminate the Universe and the Modern World
LISA RANDALL
Bodley Head: 2011.
464 pp. £20

forces with a Spanish writer in 2008 to file a lawsuit against CERN, the US Department of Energy and the US National Science Foundation that threatened to block the start-up of the LHC.

As Randall describes, scientists responded with fervour. It turns out that nature provides an answer to such concerns. Cosmic rays pervade space and bombard Earth continuously. Their energies extend to billions of times that achievable by the LHC. Had microscopic black holes been created in high-energy collisions of cosmic-ray particles, Earth and the stars would have been swallowed up long ago. Physicists could relax: the LHC risk-assessment exercise was favourably resolved.

On 20 November 2009, the LHC first powered up for experiments. By the end of 2012 it will reach a high enough energy to test the standard model of particle physics and to detect the Higgs boson, the elusive, mass-giving 'God particle' — if it exists. *Knocking on Heaven's Door* describes how that discovery would confirm one of the great predictions of physics. In parallel, the LHC will search for physics beyond the standard model. One of the most anticipated signatures will be that of supersymmetry, a new field that provides a candidate particle for dark matter.
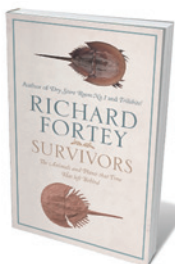
Given her background, Randall naturally complements her discussion of the LHC by describing ongoing searches for dark matter that are mostly led by particle physicists. For them, the driving question is: what is it? But Randall largely ignores astronomers' contribution to the problem — namely, giving the empirical motivation for dark matter (it is the dominant form of matter in the Universe) and mapping its location.

The LHC could resolve the greatest mysteries of the Universe: one microscopically small, and the other macroscopically large. But suppose physicists fail to detect any sign of the Higgs boson or supersymmetry? Will we have wasted those billions? Failure would shift the goal posts. Exploration of the next particle-physics frontier will require more powerful, more expensive and less attainable machines. But we would also be unsure as to how high we would need to go in terms of energy or luminosity to achieve a breakthrough in new physics. Visionary ideas would be needed.

Let us hope that the LHC does find something. And that, regardless of the outcome, the inspired efforts of its builders will combine with theorists' dreams to develop new and affordable probes of the ultimate horizons of the Universe. ∎

Joseph Silk *is professor of physics at the Institut d'Astrophysique, Université Pierre et Marie Curie, Paris, France.*
e-mail: silk@astro.ox.ac.uk

Books in brief

Survivors: The Animals and Plants that Time Has Left Behind
Richard Fortey HARPER PRESS 400 pp. £25 (2011)
Cataclysms come and go, but the stromatolites of Western Australia have sat them out for more than 2 billion years. These organic cushion-like structures with cyanobacterial wigs lead palaeontologist Richard Fortey's cast of survivors still dangling from the tree of life. He roves from hordes of horseshoe crabs in Delaware Bay on the northeast US coast to New Zealand's velvet worms and beyond, each fascinating organism a focus for broader thoughts on evolutionary history. Decades spent "looking at thoroughly dead creatures" have not dimmed Fortey's ability to bring these relics to life.

My Beautiful Genome: Discovering Our Genetic Future, One Quirk at a Time
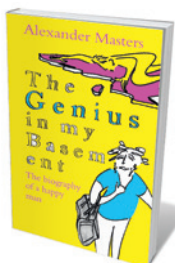*Lone Frank* ONEWORLD 320 pp. £10.99 (2011)
As consumer-led genomics ramps up, questions of ethics and efficacy proliferate. Neurobiologist Lone Frank looks at how exposing our DNA affects our lives. Having interviewed James Watson and covered the rise of personal genomics from 2008, Frank puts her own genes to the test. She charts the range of applications — deep ancestry, disease, behaviour and personality, mental illness and partner compatibility — and concludes that, far from being a straitjacket, unveiling our 'invisible self' liberates, connects and reassures.

1493: How Europe's Discovery of the Americas Revolutionized Trade, Ecology and Life on Earth
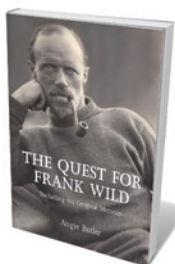*Charles C. Mann* GRANTA 544 pp. £14.99 (2011)
Journalist Charles Mann chronicles how Christopher Columbus' second New World expedition in 1493 triggered a global upheaval. European vessels left sheep, rats and lethal viruses in the New World and carried tomatoes, tobacco and maize (corn) to the Old. Millions of people died from introduced diseases and ecosystems convulsed. A world economy emerged, propelled by trade in commodities from silk to slaves. Drawing on new research, Mann reframes the past 500 years to riveting effect.

The Genius in my Basement: The Biography of a Happy Man
Alexander Masters FOURTH ESTATE 352 pp. £8.99 (2011)
In 2007, writer Alexander Masters — author of *Stuart: A Life Backwards* (2006) — lived above the distinguished mathematician Simon Phillips Norton in Cambridge, UK. Norton helped to devise the 'monstrous moonshine' conjecture, about a mathematical symmetry group in thousands of dimensions known as the Monster; he is also an eccentric who obsesses about buses and Bombay mix. Masters, with his background in maths and physics, has written a fond yet merciless portrait that attempts both to dissect the Monster and to do justice to an extraordinary character.

The Quest for Frank Wild
Angie Butler JACKLEBERRY PRESS 224 pp. £25 (2011)
Antarctic exploration is synonymous with heroes such as Ernest Shackleton, Robert Falcon Scott and Roald Amundsen. Few of us have heard of Frank Wild, Shackleton's 'right-hand man', who had pivotal roles in five Antarctic expeditions and is one of only two men to be awarded a Polar Medal with four bars. After seven years tracking Wild's fate, writer Angie Butler redresses the balance. Her account of his life includes a coup: Wild's memoir of four expeditions, including Nimrod and Endurance, is published here for the first time.

forces with a Spanish writer in 2008 to file a lawsuit against CERN, the US Department of Energy and the US National Science Foundation that threatened to block the start-up of the LHC.

As Randall describes, scientists responded with fervour. It turns out that nature provides an answer to such concerns. Cosmic rays pervade space and bombard Earth continuously. Their energies extend to billions of times that achievable by the LHC. Had microscopic black holes been created in high-energy collisions of cosmic-ray particles, Earth and the stars would have been swallowed up long ago. Physicists could relax: the LHC risk-assessment exercise was favourably resolved.

On 20 November 2009, the LHC first powered up for experiments. By the end of 2012 it will reach a high enough energy to test the standard model of particle physics and to detect the Higgs boson, the elusive, mass-giving 'God particle' — if it exists. *Knocking on Heaven's Door* describes how that discovery would confirm one of the great predictions of physics. In parallel, the LHC will search for physics beyond the standard model. One of the most anticipated signatures will be that of supersymmetry, a new field that provides a candidate particle for dark matter.

Given her background, Randall naturally complements her discussion of the LHC by describing ongoing searches for dark matter that are mostly led by particle physicists. For them, the driving question is: what is it? But Randall largely ignores astronomers' contribution to the problem — namely, giving the empirical motivation for dark matter (it is the dominant form of matter in the Universe) and mapping its location.

The LHC could resolve the greatest mysteries of the Universe: one microscopically small, and the other macroscopically large. But suppose physicists fail to detect any sign of the Higgs boson or supersymmetry? Will we have wasted those billions? Failure would shift the goal posts. Exploration of the next particle-physics frontier will require more powerful, more expensive and less attainable machines. But we would also be unsure as to how high we would need to go in terms of energy or luminosity to achieve a breakthrough in new physics. Visionary ideas would be needed.

Let us hope that the LHC does find something. And that, regardless of the outcome, the inspired efforts of its builders will combine with theorists' dreams to develop new and affordable probes of the ultimate horizons of the Universe. ■

Joseph Silk *is professor of physics at the Institut d'Astrophysique, Université Pierre et Marie Curie, Paris, France.*
e-mail: silk@astro.ox.ac.uk

Books in brief

Survivors: The Animals and Plants that Time Has Left Behind
Richard Fortey HARPER PRESS *400 pp. £25 (2011)*
Cataclysms come and go, but the stromatolites of Western Australia have sat them out for more than 2 billion years. These organic cushion-like structures with cyanobacterial wigs lead palaeontologist Richard Fortey's cast of survivors still dangling from the tree of life. He roves from hordes of horseshoe crabs in Delaware Bay on the northeast US coast to New Zealand's velvet worms and beyond, each fascinating organism a focus for broader thoughts on evolutionary history. Decades spent "looking at thoroughly dead creatures" have not dimmed Fortey's ability to bring these relics to life.

My Beautiful Genome: Discovering Our Genetic Future, One Quirk at a Time
Lone Frank ONEWORLD *320 pp. £10.99 (2011)*
As consumer-led genomics ramps up, questions of ethics and efficacy proliferate. Neurobiologist Lone Frank looks at how exposing our DNA affects our lives. Having interviewed James Watson and covered the rise of personal genomics from 2008, Frank puts her own genes to the test. She charts the range of applications — deep ancestry, disease, behaviour and personality, mental illness and partner compatibility — and concludes that, far from being a straitjacket, unveiling our 'invisible self' liberates, connects and reassures.

1493: How Europe's Discovery of the Americas Revolutionized Trade, Ecology and Life on Earth
Charles C. Mann GRANTA *544 pp. £14.99 (2011)*
Journalist Charles Mann chronicles how Christopher Columbus' second New World expedition in 1493 triggered a global upheaval. European vessels left sheep, rats and lethal viruses in the New World and carried tomatoes, tobacco and maize (corn) to the Old. Millions of people died from introduced diseases and ecosystems convulsed. A world economy emerged, propelled by trade in commodities from silk to slaves. Drawing on new research, Mann reframes the past 500 years to riveting effect.

The Genius in my Basement: The Biography of a Happy Man
Alexander Masters FOURTH ESTATE *352 pp. £8.99 (2011)*
In 2007, writer Alexander Masters — author of *Stuart: A Life Backwards* (2006) — lived above the distinguished mathematician Simon Phillips Norton in Cambridge, UK. Norton helped to devise the 'monstrous moonshine' conjecture, about a mathematical symmetry group in thousands of dimensions known as the Monster; he is also an eccentric who obsesses about buses and Bombay mix. Masters, with his background in maths and physics, has written a fond yet merciless portrait that attempts both to dissect the Monster and to do justice to an extraordinary character.

The Quest for Frank Wild
Angie Butler JACKLEBERRY PRESS *224 pp. £25 (2011)*
Antarctic exploration is synonymous with heroes such as Ernest Shackleton, Robert Falcon Scott and Roald Amundsen. Few of us have heard of Frank Wild, Shackleton's 'right-hand man', who had pivotal roles in five Antarctic expeditions and is one of only two men to be awarded a Polar Medal with four bars. After seven years tracking Wild's fate, writer Angie Butler redresses the balance. Her account of his life includes a coup: Wild's memoir of four expeditions, including Nimrod and Endurance, is published here for the first time.

Jim Ottaviani's comic-strip biography of Richard Feynman conveys the physicist's colourful personality.

Q&A Jim Ottaviani

Comic creator

Jim Ottaviani is the author of several comic books about famous scientists. His latest, with illustrations by Leland Myrick, covers the life of physicist Richard Feynman, who is known for his bongo playing and enthusiastic lectures as much as his work on quantum mechanics. Ottaviani explains why a graphic-novel format is a perfect match for such a zany character.

Why did you decide to write comic books about science?

In 1997 I was looking for comics I wanted to read. I love a good Spider-Man story, but after you have read a hundred you don't really need another. I had been a nuclear engineer, then worked as a librarian. Interested in the names behind the equations and discoveries, I read the biographies of physicist Niels Bohr and others. There were some great stories there. That got me wondering: why aren't they in comics?

Why did you choose Richard Feynman?

Every pantheon needs a trickster god. We are now far enough removed from the early foundations of quantum physics that we mythologize the people and events involved. And we like to have characters in our mythologies. Feynman slots beautifully into that. He worked hard to make his personality accessible to a broad audience, when many of his peers did not. Feynman was astute, even aggressive, about creating a myth about himself, turning his life into a sort of performance art.

Was Feynman's interest in drawing and diagrams also relevant?

With a scientific audience, sometimes you have to defend the use of images. Flip through *Physical Review*: there are a lot of pictures. We communicate science through images as well as words, so it wasn't a leap for Feynman to want to learn to draw. There are quotes from

Feynman
JIM OTTAVIANI.
ILLUSTRATED BY
LELAND MYRICK
*First Second: 2011.
272 pp. £18.99,
$29.99*

him about the beauty of a flower being accessible to both the artist and the scientist in different ways, and how they add to each other. So the comic format is valid.

Why did you decide not to present his life chronologically?

We asked: what would Feynman do? In his books, Feynman wrote in short scenes and acts, presenting some things out of order. A continuous narrative — born, lived, died — did not serve the person or the story. So using anecdotes seemed natural. Choosing the anecdotes, choosing when to break the continuity, that was when it got more difficult. Our intention was to make him seem more human.

The words and pictures in your book tell different aspects of the story. Why did you choose such a challenging format?

Leland and I assumed that the readership would be willing to read both the words and the pictures. We wanted to bring something new, to enrich the experience. Otherwise, why do another book about Feynman? We wanted to go beyond James Gleick's 1992 book *Genius* or Feynman's own stories.

What is the subject of your next book?

It is aimed at a young readership, and is about primatologists Jane Goodall, Dian Fossey and Biruté Galdikas, with a good helping of their mentor Louis Leakey. The book is targeted at readers aged between 10 and 12, mainly girls. It is nowhere near as in-depth as *Feynman*, because it is about half the length and covers three people. It is about becoming a scientist, and what it means to be one in a world that may not be prepared for you to do that sort of thing, as a woman.

Do you also touch on that issue in the Feynman book?

Yes, with Feynman's sister, Joan Feynman. We give a hint of the difficulties she faced in pursuing her career in astrophysics.

What are you trying to show in your books?

Feynman created a world in science that he enjoyed living in. Then he realized that he could have more by creating another world, where he appreciates art, plays the bongos, acts in plays or sometimes acts the fool. And this is what I have been trying to show with all the books — the humanness of science. Science is a serious endeavour. It is the old Spider-Man saying: "With great power comes great responsibility." But at the same time, scientists should appreciate the joy of it. I hope scientists will get from Feynman the idea of a full life, lived creatively, and the fun it can produce. ∎

INTERVIEW BY MARC WEIDENBAUM

Correspondence

Loophole in forest plan for Indonesia

Last year, Indonesia and Norway signed the Oslo Pact, which will pay Indonesia up to US$1 billion to reduce carbon emissions by advancing forest-conservation initiatives. As part of the deal, Indonesia must halt the licensing of new agricultural plantations and logging concessions on peatlands and natural forest for two years. Clearing and logging must instead be directed to non-forest 'degraded' lands and to existing concessions. But the pact has a big loophole.

Indonesia is the world's third-largest emitter of greenhouse gases, caused mostly by rampant felling or burning of its rainforests and carbon-rich peat-swamp forests. The loss of these ecosystems also threatens major hot spots of global biodiversity. The hope is that the Oslo Pact and follow-on carbon payments can stem this tide.

However, President Susilo Bambang Yudhoyono of Indonesia has issued a two-year moratorium on new concessions for clearing or logging of peatlands and natural primary (old-growth) forest. Contrary to the Oslo Pact, vast expanses of selectively logged forests — which sustain substantial carbon stores and much biodiversity — are classed as 'degraded' and left out of the moratorium altogether. The net effect is that these natural forests could be re-logged or cleared for oil palm and pulpwood plantations. According to its Ministry of Forestry, Indonesia has 35.4 million hectares of logged forest that can be cleared, considerably more than the upper estimate of 20 million hectares of primary forest protected under the moratorium.

Many protected forests are in steep, mountainous areas that face little threat. The most imperilled forests, in the lowlands, are largely excluded from the deal because they have been logged previously. On top of this, the moratorium fails to protect shallow peatlands from conversion, or halt primary forests and deep peatlands from being cleared for sugar cane — one of the most rapidly expanding biofuel crops.

We urge Norway to insist that logged forests and clearance for sugar cane be included under the moratorium. Without doing so, this is little more than business as usual in Indonesia.

David P. Edwards, William F. Laurance *James Cook University, Cairns, Queensland, Australia.*
dave.edwards@jcu.edu.au

Call to save science institute in Turkey

On 15 July, the Turkish Scientific and Technological Research Council (TÜBİTAK) effectively closed down the Feza Gürsey Institute for Basic Sciences in Istanbul by relocating it to a TÜBİTAK cryptology institute in Gebze. More than 1,500 signatures were collected by mid-August to ask the science minister, Nihat Ergün, to reconsider this decision.

The Feza Gürsey Institute, named after an eminent Turkish physicist, has been crucial in the training of Turkish researchers. In the words of Marta Sanz-Solé, president of the European Mathematical Society, it is central to the "consolidation of scientific international collaborations".

The institute has a remarkable research record in theoretical physics and mathematics that spans 14 years, with 350 articles published in high-profile journals and more than 2,000 citations. It has hosted international meetings and free summer schools for thousands of Turkish participants.

The move seems to be an example of TÜBİTAK's apparently low rating of basic research and its relation to applied research and technology.

Signatories in the campaign to save the institute include more than 100 prominent physicists and mathematicians, as well as the presidents of the US, European and French mathematical societies, and of Turkey's Mathematical Society, Physical Society and Astronomical Society (http://savefezagursey.wordpress.com).

Ayşe Erzan *Istanbul Technical University, Turkey.*
erzan@itu.edu.tr
Cihan Saçlıoğlu *Sabancı University, Turkey.*

Drug firm monitors waste water

At AstraZeneca we are proactively addressing the problem of pharmaceuticals entering the environment as a result of our manufacturing discharges (*Nature* **476,** 265; 2011).

Using ecotoxicity data and our knowledge of environmental fate and the local environment, we have identified long- and short-term concentrations of active pharmaceutical ingredients that we refer to as Environmental Reference Concentrations (ERCs) and Maximum Tolerable Concentrations (MTCs), respectively. These should not be exceeded in the aquatic environments that receive effluent from our manufacturing sites.

This approach is based on established environmental quality standards used in national and international legislation. Under this voluntary initiative, we have so far established ERCs and MTCs for 30 of our active pharmaceutical ingredients to protect the freshwater environment (algae, invertebrates and fish), top predators (fish-eating mammals such as otters), the marine environment (for coastal discharges) and humans. Other research-based pharmaceutical industries also have voluntary initiatives to control their discharges.

We have 'ecopharmacovigilance' procedures in place to ensure that our ERCs continue to take into account all relevant data and current scientific understanding of the fate and effects of pharmaceuticals in the environment. For example, if a new, lower, no-effect concentration is reported in the peer-reviewed literature and is scientifically robust, we revise our ERCs and environmental risk assessments accordingly.

Last year, we started a programme to monitor our own emissions against ERC and MTC values for our worldwide sites that could discharge waste water containing active pharmaceutical ingredients during peak production. We are now starting to share the ERC approach with our third-party manufacturers, with a view to including them in the programme.

Jason Snape, Chris Lewis, Richard Murray-Smith *AstraZeneca, Brixham Environmental Laboratory, UK.*
richard.murray-smith@astrazeneca.com
Competing financial interests declared. See go.nature.com/e6tkkb.

Diploma database to encourage mobility

Researchers moving abroad often need to have their qualifications recognized by a local university or other national institution. This costly process can take months, and may include thesis re-evaluation by a panel of professors or researchers. Add to this the extensive paperwork already required for applications for jobs, research grants or scholarships, and the need to cut red tape to encourage mobility within the scientific community becomes more pressing.

Governments and universities should create an international online database that details the qualifications of applicants from foreign universities, together with a summary of the work entailed and the standard expected.

Decisions by institutions responsible for recognizing diplomas awarded abroad would then be just a few clicks away.

Ana M. C. Santos *Federal University of Goiás, Goiânia, Brazil. ana.margarida.c.santos@googlemail.com*

Baruj Benacerraf
(1920–2011)

Immunologist who won Nobel for genetics of T-cell antigen recognition.

Baruj Benacerraf bestrode immunology for five decades. He created the intellectual framework leading to our present understanding of how T lymphocytes recognize antigens, for which he received the 1980 Nobel Prize in Physiology or Medicine. He crafted world-renowned centres of immunology at multiple institutions, oversaw the flowering of a premier cancer centre, and was a remarkable mentor to generations of immunologists. He died, aged 90, on 2 August 2011.

Born in 1920 in Caracas, Venezuela, to Sephardic Jewish parents, Benacerraf spent his youth in Paris and emigrated to the United States in 1940. He did his undergraduate studies as a premedical student at Columbia University in New York, where he met Annette Dreyfus, also a French Jewish émigré. They married in 1943, a love match that ended only in June this year with Annette's death. She travelled everywhere with him and was a constant presence in the lab, softening his rough edges, reminding him to take his medicines, pampering him with his favourite French cookies and tea.

After completing his MD at the Medical College of Virginia in Richmond, Benacerraf served in the US Army in France and returned to New York City in 1947. He began his scientific career in 1948 at Columbia with Elvin Kabat, a leading figure in immunochemistry. Family business responsibilities took him to Paris in 1949 where, at the Broussais Hospital, he carried out groundbreaking studies on the 'reticuloendothelial system', then the term for the system of white blood cells (phagocytes) that ingest foreign particles and cell debris. Recruited to New York University (NYU) in 1956, he became the intellectual centre of a distinguished group of scientists and began the training activities that were to be one of the great achievements of his life.

At NYU, Benacerraf proved to be an immunological polymath, uncovering the existence and diverse functions of antibody subclasses, identifying cellular receptors for these proteins and establishing the distinctive antigen-recognition properties of B and T lymphocytes. He also continued the theme of combining business and scientific responsibilities, serving as a bank director for most of his NYU tenure.

His great work was the discovery that in a population of outbred guinea pigs, immune responses to simple antigens could be mounted by some animals but not by others. He showed that responsiveness was controlled by a genetic locus that determined whether the immune system could perceive the material and generate a functional response — a fundamental insight that led to his Nobel prize. These immune response (Ir) genes were later shown to be linked to the major histocompatibility complex (MHC) — which contains many immunity-related genes — and were eventually determined to code for the MHC molecules themselves.
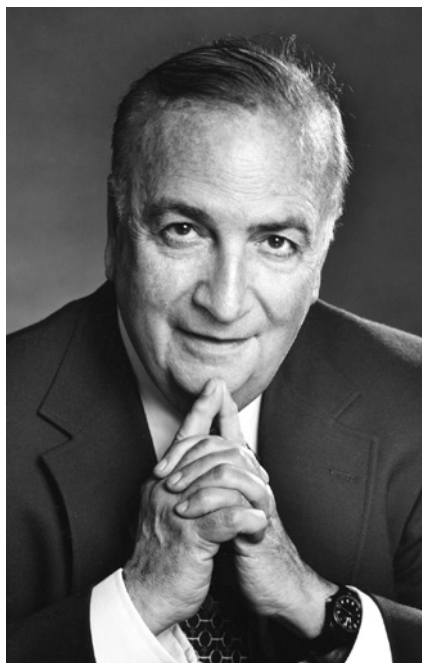
Benacerraf left NYU in 1968 and led the National Institute of Allergy and Infectious Diseases's Laboratory of Immunology in Bethesda, Maryland, for two years, where his studies led to a fuller understanding of how Ir gene products mediate immunity. He also took the first steps that led to the laboratory becoming a major centre of immunology research. In 1970, he became chair of the pathology department at Harvard Medical School in Boston, Massachusetts, a position he held until 1991.

On his arrival, the faculty included some immunological luminaries, but the school was not the world-leading centre of immunology it is today. Benacerraf recruited outstanding immunologists to various Harvard hospitals and centres in addition to the 'Quadrangle' faculty he led. His own efforts to understand the significance of T-cell recognition of peptide MHC complexes flourished and he continued his extraordinary record of training young scientists who later would become leaders in the field. In 1980, Benacerraf was appointed president of the Sidney Farber Cancer Institute in Boston, now the Dana–Farber Cancer Institute. His leadership transformed it into one of the premier cancer centres and medical research organizations in the world. He stepped down from the presidency in 1992.

Those of us who passed through his laboratory at Harvard were invited to spend idyllic summer days at Woods Hole in the time before e-mail and the web. We shared ocean breezes and excellent food with Baruj and Annette and had intense scientific discussions that he expected to be translated into concrete experiments immediately on our return.

A rare combination of capacities made Benacerraf such a successful scientist and leader. His joy at uncovering some new aspect of the immune system was almost childlike: his impish smile, finger snapping and jigging footwork at 'Aha!' moments were well known to his colleagues and trainees. He had enormous native intelligence, remarkable scientific intuition and a tremendous capacity to recognize latent talent, as well as to inspire, motivate and guide the careers of those in whom he identified such potential. It was sometimes uncomfortable in the moment when one's 'buttons' were being pushed, but in retrospect the value of such productive manipulation was always apparent.

Indeed, Benacerraf was a man of enormous personal warmth — an attribute not always appreciated by those outside his close scientific family. As pleased as he was with his achievements in science and scientific administration, he took every opportunity to remind those close to him that his main pride was in the cadre of scientists he trained. Their continuing scientific achievements are his greatest legacy. ■

Ronald N. Germain and **William E. Paul** *are at the Laboratory of Systems Biology and Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. RNG worked closely with Benacerraf from 1972 to 1981; WEP did so from 1964 to 1970.*
e-mails: rgermain@nih.gov; wpaul@niaid.nih.gov

Y. KARSH/CAMERA PRESS LONDON

NEWS & VIEWS

How to escape treatment

Even during effective treatment with antiretroviral drugs, low levels of HIV persist. In part, this could be due to cell–to–cell transfer of multiple virions and the drugs' inability to inhibit replication when virus levels are high. SEE LETTER P.95

STEVEN G. DEEKS

One of the great triumphs of modern medicine is the development of combination antiretroviral drug therapy for the management of HIV infection. For those with access to these drugs, modern regimens can reduce the amount of circulating virus to very low levels. Despite their inherent potency, however, antiretroviral drugs are not curative, and HIV persists indefinitely. Consequently, patients have to adhere to these expensive and potentially toxic drugs for life. On page 95 of this issue, Sigal and colleagues[1] provide insights into why antiretroviral drugs cannot fully inhibit HIV replication. Their data have notable implications for the emerging efforts aimed at curing HIV infection[2].

Several mechanisms might contribute to HIV persistence during antiretroviral therapy. These include maintenance of a transcriptionally silent (latent) HIV genome in long-lived, resting target cells such as CD4[+] memory T cells; proliferation of these latently infected cells; inadequate anti-HIV clearance mechanisms; and ongoing *de novo* infection of susceptible target cells through continued viral replication[3,4]. Although most researchers agree on the involvement of the first three mechanisms, the degree to which HIV can effectively replicate during therapy is a highly contentious issue.

The arguments against persistent replication are that the virus does not evolve in peripheral blood, and that following intensification — the process whereby potent drugs are added to a stable regimen — the steady-state levels of HIV RNA in the plasma do not change[5].

Nevertheless, several lines of evidence support ongoing replication. Preliminary studies[6,7] suggest that intensification may affect levels of the virus in tissues known to be enriched in HIV-susceptible target cells; the rapid reduction in virus levels in response to an antiretroviral drug can most easily be explained by the inhibition of the ongoing complete cycles of replication. Also, cells of many treated individuals contain unintegrated episomal HIV DNA (a potential marker of recent cell infection)[8]. Finally, activated cells contain higher levels of HIV DNA than resting cells do[9] — an observation that is more readily explained by
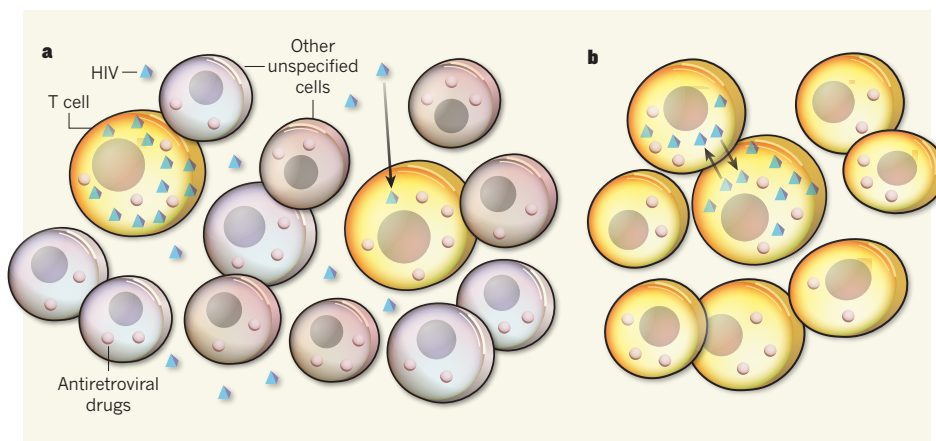
Figure 1 | HIV infection of target cells. HIV infects CD4[+] memory T cells, which are scattered throughout the body. **a**, Where these cells are sparse (for instance, at effector sites in the mucosa) successful infection is likely to involve cell-free transfer of single virions (blue pyramids) to distant target cells (arrow). **b**, But in areas of high T-cell density (such as inductive tissues in lymph nodes), transmission is likely to involve direct transfer of multiple virions into adjacent cells. Sigal and colleagues' findings[1] suggest that antiretroviral drugs (small pink circles) readily inhibit cell-free transfer of single virions, but may not completely stop cell-to-cell transfer of virions in target-rich areas.

active cycles of infection than by preferential activation of cells that harbour latent virus.

Sigal *et al.*[1] provide a compelling and intuitive mechanism that might account for the failure of potent drugs to completely inhibit transfer of replication-competent virus. Using a complex mathematical model, they show that the drug concentration required to prevent a single transmitted virion from successfully infecting a target cell is much lower than that needed to stop multiple transmitted virus particles from infecting the same cell. They then show in an *in vitro* system that physiologically relevant drug concentrations can readily inhibit cell-free transmission — that is, transmission of a free virus to a target cell — but not cell-to-cell transfer of the virus (Fig. 1). These modelling and experimental findings are consistent, because cell-to-cell transmission is known[10,11] to be associated with the transfer of multiple virions to the target cell.

Sigal and co-workers further show that under conditions of maximal drug exposure, the replication ratio following cell-to-cell transfer was less than 1, yet greater than that predicted for complete suppression. This suggests that although replication may not fully account for HIV persistence, it is likely to be a

contributing factor. Finally, the authors argue through another mathematical model that low-level replication of the virus can replenish the virus reservoir even in the absence of obvious evolution, because localized tissue-based chains of infection are independent, intermittent and unlikely to be linked temporally. This is yet another intuitive conclusion that is easy to place in the context of existing evidence, but like all the predictions in this paper it is exceedingly difficult to prove *in vivo*.

I should emphasize that these observations[1] are not definitive. An important matter for the field is to define whether cell-to-cell virus transmission does indeed occur during effective therapy. Given the logistics of accessing lymphoid tissues and their T cells in humans, it may be necessary to develop a robust non-human primate model in which animals are treated for prolonged periods. Such work could complement the ongoing studies aimed at defining the size and distribution of the HIV reservoir in humans.

One might argue that the existing antiretroviral regimens will probably be effective as long as a patient harbours drug-susceptible virus and adheres to the correct dosage of the drugs. So why should we care whether there is

a small amount of difficult-to-prove *de novo* infection of certain cells in lymphoid tissues?

There is growing recognition that delivering antiretroviral drugs for life to everyone who might benefit from them is not going to be possible. Despite the unprecedented global investment in providing drugs, the number of newly diagnosed infections remains far greater than the number of people with access to therapy. As resources are likely to become more limited over time, we will not be able to treat our way out of this epidemic. And even if therapy is delivered, many people cannot fully adhere to it over timescales of years to decades. These public-health and individual failures result in significant harm to the individuals (owing to untreated progressive HIV infection) and to the society (because untreated people are far more likely than treated people to transmit the virus to others[12]).

The only way to fully address individual and public needs is to cure those infected with HIV. Given the existence of long-lived reservoirs of HIV, even complete inhibition of this virus's replication is unlikely to clear it completely from the body. But achieving complete inhibition is almost certainly going to be necessary for the

many other curative strategies now being considered[2]. Addressing the questions Sigal and colleagues[1] raise in a more definitive *in vivo* experiment will be a challenge, but it is probably among the most crucial tasks for those who conduct translational research in this area. ∎

Steven G. Deeks *is in the Department of Medicine, University of California, San Francisco, San Francisco, California 94110, USA.*
e-mail: sdeeks@php.ucsf.edu

1. Sigal, A. *et al. Nature* **477,** 95–98 (2011).
2. Cohen, J. *Science* **332,** 784–789 (2011).
3. Richman, D. D. *et al. Science* **323,** 1304–1307 (2009).
4. Trono, D. *et al. Science* **329,** 174–180 (2010).
5. Dinoso, J. B. *et al. Proc. Natl Acad. Sci. USA* **106,** 9403–9408 (2009).
6. Buzón, M. J. *et al. Nature Med.* **16,** 460–465 (2010).
7. Yukl, S. A. *et al. AIDS* **24,** 2451–2460 (2010).
8. Sharkey, M. *et al. PLoS Pathog.* **7,** e1001303 (2011).
9. Chun, T. W. *et al. J. Clin. Invest.* **115,** 3250–3255 (2005).
10. Jung, A. *et al. Nature* **418,** 144 (2002).
11. Del Portillo, A. *et al. J. Virol.* **85,** 7169–7176 (2011).
12. Cohen, M. S. *et al. N. Engl. J. Med.* **365,** 493–505 (2011).

ASTROPHYSICS

A hint of normality at last?

The chemical diversity of the oldest stars is greater than we thought. The discovery of an extremely iron-poor star with a 'normal' ratio of carbon to iron challenges our perception of early chemical enrichment. SEE LETTER P.67

JOHN E. NORRIS

The chemical abundances of the oldest stars in the Galaxy hold a key to what the Universe's conditions were like at the earliest times. Indeed, the three most chemically primitive stars currently known have abundance patterns extremely different from those formed after the Universe's first billion years — for example, their abundance ratio of carbon to iron is about 10–1,000 times larger than is found at later times, indicating quite different conditions between the two epochs. On page 67 of this issue, Caffau *et al.*[1] report a chemically primitive star that is in some respects more 'normal' than these objects, indicating a larger chemical diversity, which challenges our understanding of the first stars.

Most astrophysicists agree that the Big Bang hypothesis provides the best description of the formation of the Universe. According to the standard version of this theory, a few minutes after the Universe began the only chemical

elements were hydrogen, helium and lithium. At that time, some 13.7 billion years ago, their fractions by mass were 0.75, 0.25 and 2.8×10^{-9}, respectively[2,3]. Astronomers, on the other hand, have observed no stars devoid of elements more massive than lithium. The two most iron-poor stars, which are believed to have ages of approximately 13 billion years, have an observed iron abundance about $10^{-5.5}$ that of the Sun — a small but well-determined amount. Further, the observed fractional lithium abundance in most stars that formed close in time to the beginning of the Universe, and in which the observed abundances should not have changed from their initial values, is 8.3×10^{-10}, some three times smaller than predicted[4]. Finally, in the three most iron-poor stars (all with less than $10^{-4.5}$ solar iron abundance), carbon, nitrogen and oxygen are present in prodigious amounts relative to iron; and in one of them, the abundances of sodium, magnesium and aluminium relative to iron are at least 100 times those of the Sun.

What do these observed abundances tell us?

One suggestion is that supernovae (the final explosions of massive stars, which produce essentially all of the elements heavier than lithium and enrich the gas clouds from which later stars form) were very different within the first few hundred million years from those that followed, leading to the very different abundance patterns. Another possibility is that large overabundances of carbon and/or oxygen may have had a crucial role in determining the nature of the first stars to form.

A hint of normality has been restored to the field by Caffau and colleagues' discovery[1] of the extremely iron-poor dwarf star, SDSS J102915+172927, which has $10^{-5.0}$ the iron abundance of the Sun. This value places it in the same range as the three most iron-poor stars mentioned above. Normality prevails in the sense that this star is not strongly carbon enhanced: the authors did not detect carbon in it, and the carbon-to-iron abundance ratio upper limit is not too different from the solar ratio. As Caffau *et al.* point out, the low carbon abundance in this object seems inconsistent with the prediction of Frebel *et al.*[5] that large carbon and/or oxygen abundances is an essential ingredient that provides cooling of the gas clouds from which the early low-mass, long-lived stars we observe today were formed.

The question that begs to be answered is: what does 'normal' mean during the Universe's first few hundred million years? Given that the majority of these four most iron-poor stars is carbon-rich, should the carbon-rich stars not be considered normal and the 'carbon-normal' object abnormal? More to the point, should one think in terms of two different types of chemical-enrichment sources that produce the different observed chemical signatures — perhaps the 'mixing and fallback' type of supernova[6] for the carbon-rich stars and the standard 'core-collapse' supernovae for the carbon-normal — or does one need something quite different? Another intriguing question is whether the 'carbon-normal' SDSS J102915+172927 is more primitive than the three carbon-rich members of the most iron-poor stars discussed here.

There will be considerable interest in the lithium abundance of this star, which is anything but normal: Caffau *et al.*[1] were unable to detect lithium in the spectrum of SDSS J102915+172927 (despite the expectation that it should be readily detectable) and report a lithium mass fraction of less than 6.8×10^{-11} — more than 40 times smaller than the predictions of Big Bang nucleosynthesis, the process by which atomic nuclei were formed in the early Universe. Only one of the previously known three most iron-poor stars (HE 1327-2326) has an effective temperature at which stellar evolutionary effects are not expected to have greatly altered its original lithium abundance, and this star is also lithium deficient[7], by a factor greater than 100. Thus, as far as we know, all four could have been born

a small amount of difficult-to-prove *de novo* infection of certain cells in lymphoid tissues?

There is growing recognition that delivering antiretroviral drugs for life to everyone who might benefit from them is not going to be possible. Despite the unprecedented global investment in providing drugs, the number of newly diagnosed infections remains far greater than the number of people with access to therapy. As resources are likely to become more limited over time, we will not be able to treat our way out of this epidemic. And even if therapy is delivered, many people cannot fully adhere to it over time-scales of years to decades. These public-health and individual failures result in significant harm to the individuals (owing to untreated progressive HIV infection) and to the society (because untreated people are far more likely than treated people to transmit the virus to others[12]).

The only way to fully address individual and public needs is to cure those infected with HIV. Given the existence of long-lived reservoirs of HIV, even complete inhibition of this virus's replication is unlikely to clear it completely from the body. But achieving complete inhibition is almost certainly going to be necessary for the

many other curative strategies now being considered[2]. Addressing the questions Sigal and colleagues[1] raise in a more definitive *in vivo* experiment will be a challenge, but it is probably among the most crucial tasks for those who conduct translational research in this area. ∎

Steven G. Deeks *is in the Department of Medicine, University of California, San Francisco, San Francisco, California 94110, USA.*
e-mail: sdeeks@php.ucsf.edu

1. Sigal, A. *et al. Nature* **477,** 95–98 (2011).
2. Cohen, J. *Science* **332,** 784–789 (2011).
3. Richman, D. D. *et al. Science* **323,** 1304–1307 (2009).
4. Trono, D. *et al. Science* **329,** 174–180 (2010).
5. Dinoso, J. B. *et al. Proc. Natl Acad. Sci. USA* **106,** 9403–9408 (2009).
6. Buzón, M. J. *et al. Nature Med.* **16,** 460–465 (2010).
7. Yukl, S. A. *et al. AIDS* **24,** 2451–2460 (2010).
8. Sharkey, M. *et al. PLoS Pathog.* **7,** e1001303 (2011).
9. Chun, T. W. *et al. J. Clin. Invest.* **115,** 3250–3255 (2005).
10. Jung, A. *et al. Nature* **418,** 144 (2002).
11. Del Portillo, A. *et al. J. Virol.* **85,** 7169–7176 (2011).
12. Cohen, M. S. *et al. N. Engl. J. Med.* **365,** 493–505 (2011).

ASTROPHYSICS

A hint of normality at last?

The chemical diversity of the oldest stars is greater than we thought. The discovery of an extremely iron-poor star with a 'normal' ratio of carbon to iron challenges our perception of early chemical enrichment. SEE LETTER P.67

JOHN E. NORRIS

The chemical abundances of the oldest stars in the Galaxy hold a key to what the Universe's conditions were like at the earliest times. Indeed, the three most chemically primitive stars currently known have abundance patterns extremely different from those formed after the Universe's first billion years — for example, their abundance ratio of carbon to iron is about 10–1,000 times larger than is found at later times, indicating quite different conditions between the two epochs. On page 67 of this issue, Caffau *et al.*[1] report a chemically primitive star that is in some respects more 'normal' than these objects, indicating a larger chemical diversity, which challenges our understanding of the first stars.

Most astrophysicists agree that the Big Bang hypothesis provides the best description of the formation of the Universe. According to the standard version of this theory, a few minutes after the Universe began the only chemical

elements were hydrogen, helium and lithium. At that time, some 13.7 billion years ago, their fractions by mass were 0.75, 0.25 and 2.8×10^{-9}, respectively[2,3]. Astronomers, on the other hand, have observed no stars devoid of elements more massive than lithium. The two most iron-poor stars, which are believed to have ages of approximately 13 billion years, have an observed iron abundance about $10^{-5.5}$ that of the Sun — a small but well-determined amount. Further, the observed fractional lithium abundance in most stars that formed close in time to the beginning of the Universe, and in which the observed abundances should not have changed from their initial values, is 8.3×10^{-10}, some three times smaller than predicted[4]. Finally, in the three most iron-poor stars (all with less than $10^{-4.5}$ solar iron abundance), carbon, nitrogen and oxygen are present in prodigious amounts relative to iron; and in one of them, the abundances of sodium, magnesium and aluminium relative to iron are at least 100 times those of the Sun.

What do these observed abundances tell us?

One suggestion is that supernovae (the final explosions of massive stars, which produce essentially all of the elements heavier than lithium and enrich the gas clouds from which later stars form) were very different within the first few hundred million years from those that followed, leading to the very different abundance patterns. Another possibility is that large overabundances of carbon and/or oxygen may have had a crucial role in determining the nature of the first stars to form.

A hint of normality has been restored to the field by Caffau and colleagues' discovery[1] of the extremely iron-poor dwarf star, SDSS J102915+172927, which has $10^{-5.0}$ the iron abundance of the Sun. This value places it in the same range as the three most iron-poor stars mentioned above. Normality prevails in the sense that this star is not strongly carbon enhanced: the authors did not detect carbon in it, and the carbon-to-iron abundance ratio upper limit is not too different from the solar ratio. As Caffau *et al.* point out, the low carbon abundance in this object seems inconsistent with the prediction of Frebel *et al.*[5] that large carbon and/or oxygen abundances is an essential ingredient that provides cooling of the gas clouds from which the early low-mass, long-lived stars we observe today were formed.

The question that begs to be answered is: what does 'normal' mean during the Universe's first few hundred million years? Given that the majority of these four most iron-poor stars is carbon-rich, should the carbon-rich stars not be considered normal and the 'carbon-normal' object abnormal? More to the point, should one think in terms of two different types of chemical-enrichment sources that produce the different observed chemical signatures — perhaps the 'mixing and fallback' type of supernova[6] for the carbon-rich stars and the standard 'core-collapse' supernovae for the carbon-normal — or does one need something quite different? Another intriguing question is whether the 'carbon-normal' SDSS J102915+172927 is more primitive than the three carbon-rich members of the most iron-poor stars discussed here.

There will be considerable interest in the lithium abundance of this star, which is anything but normal: Caffau *et al.*[1] were unable to detect lithium in the spectrum of SDSS J102915+172927 (despite the expectation that it should be readily detectable) and report a lithium mass fraction of less than 6.8×10^{-11} — more than 40 times smaller than the predictions of Big Bang nucleosynthesis, the process by which atomic nuclei were formed in the early Universe. Only one of the previously known three most iron-poor stars (HE 1327-2326) has an effective temperature at which stellar evolutionary effects are not expected to have greatly altered its original lithium abundance, and this star is also lithium deficient[7], by a factor greater than 100. Thus, as far as we know, all four could have been born

lithium-poor. Although lithium-deficient stars are not unknown in the Galaxy's halo, which contains the oldest and more iron-poor stars in the system and in which the lithium deficiency is driven perhaps by their being members of binary systems[8], they comprise only about 5% of the Galaxy's halo stars. The absence of lithium in the most iron-poor stars discussed here is an exciting and potentially fundamental result. What has happened to the lithium created at the birth of the Universe?

The caveat to the above discussion is, of course, the small number of currently known iron-poor stars that have less than $10^{-4.5}$ the solar iron abundance. Caffau et al.[1] comment that they expect 5–50 stars of similar (or lower) iron content to that of SDSS J102915+172927 to be found in the Sloan Digital Sky Survey, in which they discovered this star. If they, and other currently planned surveys for the most metal-poor stars, are successful, the long-standing tyranny of small numbers will indeed have been overcome. ■

John E. Norris is in the Research School of Astronomy and Astrophysics, The Australian National University, Canberra, ACT 0200, Australia.
e-mail: jen@mso.anu.edu.au

1. Caffau, E. et al. Nature **477,** 67–69 (2011).
2. Spergel, D. N. et al. Astrophys. J. Suppl. **170,** 377–408 (2007).
3. Cyburt, R. H., Fields, B. D. & Olive, K. A. J. Cosmol. Astro-Part. Phys. **11,** 012 (2008).
4. Sbordone, L. et al. Astron. Astrophys. **522,** A26 (2010).
5. Frebel, A., Johnson, J. L. & Bromm, V. Mon. Not. R. Astron. Soc. **380,** L40–L44 (2007).
6. Umeda, H. & Nomoto, K. Nature **422,** 871–873 (2003).
7. Frebel, A., Collet, R., Eriksson, K., Christlieb, N. & Aoki, W. Astrophys. J. **684,** 588–602 (2008).
8. Ryan, S. G., Gregory, S. G., Kolb, U., Beers, T. C. & Kajino, T. Astrophys. J. **571,** 501–511 (2002).

ECOLOGY

Nitrogen from the deep

Ecosystems acquire nitrogen from the atmosphere, but this source can't account for the large nitrogen capital of some systems. The finding that bedrock can also act as a nitrogen source may help solve the riddle. SEE LETTER P.78

EDWARD A. G. SCHUUR

In many parts of the world, the rates at which plants grow are controlled by the availability of nitrogen, an essential element for all life[1]. Forest growth is fuelled by the nitrogen capital contained in soils and biomass. Like money in a bank account, this capital increases slowly within ecosystems over hundreds to thousands of years from the accumulation of tiny deposits of nitrogen that arrive each year from the atmosphere[2]. But on page 78 of this issue, Morford et al.[3] report that nitrogen-rich bedrock can double nitrogen input rates to forest ecosystems, which flourish as a result.

Nitrogen is the fourth most abundant element in living organisms, and is used as a building block for critically important biological molecules such as amino acids and nucleic acids. In many ecosystems worldwide, nitrogen is the element whose supply rate from the environment is most limited. Because competition is fierce for this resource, nitrogen supply controls the behaviour of many organisms and shapes the structure and function of whole ecosystems.

Most of the nitrogen needed by organisms to grow is supplied by recycling, in which decomposing organic matter releases nitrogen in forms that can be acquired by plants and microorganisms. Recycling, in turn, is dependent on the nitrogen capital that has accumulated over time in an ecosystem. New inorganic nitrogen is deposited into ecosystems abiotically in rainfall, or with the assistance of certain microbes that, individually or in close relationships with plants or fungi, convert inert atmospheric nitrogen gas into a form that organisms can use.

The new deposits of nitrogen are small relative to the quantities of the element that are recycled. But they are vital, not only as a source of nitrogen for newly forming ecosystems, but also because they sustain growth over centuries of ecosystem development by balancing natural nitrogen loss out of ecosystems into streams or back to the atmosphere. Scientists have made detailed measurements of atmospheric nitrogen inputs in many places, but have sometimes encountered a puzzling phenomenon: the nitrogen capital within some ecosystems is larger than can be accounted for by known atmospheric sources[4].

Morford and colleagues' discovery[3] that bedrock can provide substantial quantities of nitrogen to organisms provides a new piece of the puzzle, and in doing so helps reshape our view of ecosystem nitrogen budgets. The authors compared plants and soils from forest ecosystems in California that are similar in terms of their stand age, climate and soil type, but which grow on two different types of bedrock. They found that tree species common to both sites were enriched in nitrogen in forests growing on soils derived from mica-schist (a type of marine sedimentary rock) compared with those in forests growing on soils derived from gabbro-diorite (a type of igneous rock). Tellingly, the sedimentary rock contained roughly ten times the levels of nitrogen found in the igneous rock.

The authors found that the trees growing on the sedimentary rock not only had higher nitrogen levels, but also had more leaves than did trees growing on the igneous rock (Fig. 1). This presumably enables them to grow faster and results in a more productive forest. The elevated nitrogen levels in these trees corresponded to the measured nitrogen capital of the underlying soils, which are the direct source of the nitrogen to the forest — that is, the nitrogen capital in these soils was twice that of conifer-forest soils overlying igneous rock.

Nitrogen levels vary across forest stands for many reasons, and so Morford et al. needed more evidence to show that bedrock was responsible for the observed variations. The authors therefore took advantage of the fact that plants, soils and bedrock all

Figure 1 | Flourishing forests. Morford et al.[3] report that conifer forests at South Fork Mountain, California, are enriched in nitrogen supplied by the underlying sedimentary rock. This nitrogen boost increases the above-ground biomass of the forest.

S. MORFORD

lithium-poor. Although lithium-deficient stars are not unknown in the Galaxy's halo, which contains the oldest and more iron-poor stars in the system and in which the lithium deficiency is driven perhaps by their being members of binary systems[8], they comprise only about 5% of the Galaxy's halo stars. The absence of lithium in the most iron-poor stars discussed here is an exciting and potentially fundamental result. What has happened to the lithium created at the birth of the Universe?

The caveat to the above discussion is, of course, the small number of currently known iron-poor stars that have less than $10^{-4.5}$ the solar iron abundance. Caffau et al.[1] comment that they expect 5–50 stars of similar (or lower) iron content to that of SDSS J102915+172927 to be found in the Sloan Digital Sky Survey, in which they discovered this star. If they, and other currently planned surveys for the most metal-poor stars, are successful, the long-standing tyranny of small numbers will indeed have been overcome. ∎

John E. Norris *is in the Research School of Astronomy and Astrophysics, The Australian National University, Canberra, ACT 0200, Australia.*
e-mail: jen@mso.anu.edu.au

1. Caffau, E. et al. Nature **477,** 67–69 (2011).
2. Spergel, D. N. et al. Astrophys. J. Suppl. **170,** 377–408 (2007).
3. Cyburt, R. H., Fields, B. D. & Olive, K. A. J. Cosmol. Astro-Part. Phys. **11,** 012 (2008).
4. Sbordone, L. et al. Astron. Astrophys. **522,** A26 (2010).
5. Frebel, A., Johnson, J. L. & Bromm, V. Mon. Not. R. Astron. Soc. **380,** L40–L44 (2007).
6. Umeda, H. & Nomoto, K. Nature **422,** 871–873 (2003).
7. Frebel, A., Collet, R., Eriksson, K., Christlieb, N. & Aoki, W. Astrophys. J. **684,** 588–602 (2008).
8. Ryan, S. G., Gregory, S. G., Kolb, U., Beers, T. C. & Kajino, T. Astrophys. J. **571,** 501–511 (2002).

ECOLOGY

Nitrogen from the deep

Ecosystems acquire nitrogen from the atmosphere, but this source can't account for the large nitrogen capital of some systems. The finding that bedrock can also act as a nitrogen source may help solve the riddle. SEE LETTER P.78

EDWARD A. G. SCHUUR

In many parts of the world, the rates at which plants grow are controlled by the availability of nitrogen, an essential element for all life[1]. Forest growth is fuelled by the nitrogen capital contained in soils and biomass. Like money in a bank account, this capital increases slowly within ecosystems over hundreds to thousands of years from the accumulation of tiny deposits of nitrogen that arrive each year from the atmosphere[2]. But on page 78 of this issue, Morford et al.[3] report that nitrogen-rich bedrock can double nitrogen input rates to forest ecosystems, which flourish as a result.

Nitrogen is the fourth most abundant element in living organisms, and is used as a building block for critically important biological molecules such as amino acids and nucleic acids. In many ecosystems worldwide, nitrogen is the element whose supply rate from the environment is most limited. Because competition is fierce for this resource, nitrogen supply controls the behaviour of many organisms and shapes the structure and function of whole ecosystems.

Most of the nitrogen needed by organisms to grow is supplied by recycling, in which decomposing organic matter releases nitrogen in forms that can be acquired by plants and microorganisms. Recycling, in turn, is dependent on the nitrogen capital that has accumulated over time in an ecosystem. New inorganic nitrogen is deposited into ecosystems abiotically in rainfall, or with the assistance of certain microbes that, individually or in close relationships with plants or fungi, convert inert atmospheric nitrogen gas into a form that organisms can use.

The new deposits of nitrogen are small relative to the quantities of the element that are recycled. But they are vital, not only as a source of nitrogen for newly forming ecosystems, but also because they sustain growth over centuries of ecosystem development by balancing natural nitrogen loss out of ecosystems into streams or back to the atmosphere. Scientists have made detailed measurements of atmospheric nitrogen inputs in many places, but have sometimes encountered a puzzling phenomenon: the nitrogen capital within some ecosystems is larger than can be accounted for by known atmospheric sources[4].

Morford and colleagues' discovery[3] that bedrock can provide substantial quantities of nitrogen to organisms provides a new piece of the puzzle, and in doing so helps reshape our view of ecosystem nitrogen budgets. The authors compared plants and soils from forest ecosystems in California that are similar in terms of their stand age, climate and soil type, but which grow on two different types of bedrock. They found that tree species common to both sites were enriched in nitrogen in forests growing on soils derived from mica-schist (a type of marine sedimentary rock) compared with those in forests growing on soils derived from gabbro-diorite (a type of igneous rock). Tellingly, the sedimentary rock contained roughly ten times the levels of nitrogen found in the igneous rock.

The authors found that the trees growing on the sedimentary rock not only had higher nitrogen levels, but also had more leaves than did trees growing on the igneous rock (Fig. 1). This presumably enables them to grow faster and results in a more productive forest. The elevated nitrogen levels in these trees corresponded to the measured nitrogen capital of the underlying soils, which are the direct source of the nitrogen to the forest — that is, the nitrogen capital in these soils was twice that of conifer-forest soils overlying igneous rock.

Nitrogen levels vary across forest stands for many reasons, and so Morford et al. needed more evidence to show that bedrock was responsible for the observed variations. The authors therefore took advantage of the fact that plants, soils and bedrock all

Figure 1 | Flourishing forests. Morford et al.[3] report that conifer forests at South Fork Mountain, California, are enriched in nitrogen supplied by the underlying sedimentary rock. This nitrogen boost increases the above-ground biomass of the forest.

S. MORFORD

contain measurably different amounts of ^{15}N in their nitrogen pools. They found that, in the forest growing on nitrogen-rich sedimentary rock, the ^{15}N-content in both plants and soils matched that of the bedrock; this was not true for forests growing on the nitrogen-poor igneous rock, ruling out the possibility of significant nitrogen contribution from this rock.

Although the nitrogen-isotope measurements helped build the case for sedimentary bedrock as a nitrogen source for forests, they alone were not a smoking gun. To extend the findings beyond the carefully matched forest stands, the authors carried out a regional analysis of similar conifer forests in California. Sure enough, they found that the above-ground biomass of forests growing on nitrogen-rich sedimentary bedrock was almost 50% bigger by mass than that of forests on igneous bedrock,

after accounting for differing ages of tree stands.

The 'imprint' of nitrogen from bedrock on streams[5] and soils[6] has previously been reported for isolated sites in the same general region as the current study[3], and so Morford and colleagues' analysis makes the case for this as a regional pattern. But less than 2% of conifer-forest soils in that same region have a nitrogen capital as high as the sedimentary-bedrock forest that has been intensively studied by the authors (see Supplementary Information for ref. 3). This means that the high input of nitrogen from bedrock beneath that forest — which is equivalent to atmospheric nitrogen inputs — probably represents an upper estimate for the extent of this phenomenon. With 75% of Earth covered by sedimentary and related rock types[7], there is a real need to explore the phenomenon beyond this region to determine

what more common levels of bedrock-nitrogen inputs are for ecosystems elsewhere. ∎

Edward A. G. Schuur *is in the Department of Biology, University of Florida, Gainesville, Florida 32611-8526, USA.*
e-mail: tschuur@ufl.edu

1. Vitousek, P. M. & Howarth, R. W. *Biogeochemistry* **13,** 87–115 (1991).
2. Vitousek, P. M. *Nutrient Cycling and Limitation: Hawai'i as a Model System* (Princeton Univ. Press, 2004).
3. Morford, S. L., Houlton, B. Z. & Dahlgren, R. A. *Nature* **477,** 78–81 (2011).
4. Binkley, D., Son, Y. & Valentine, D. W. *Ecosystems* **3,** 321–331 (2000).
5. Holloway, J. M. *et al. Nature* **395,** 785–788 (1998).
6. Dahlgren, R. A. *Nature* **368,** 838–841 (1994).
7. Blatt, H. & Jones, R. L. *Geol. Soc. Am. Bull.* **86,** 1085–1088 (1975).

CANCER

Tumour–fighting virus homes in

An early clinical trial demonstrates the delivery and replication of a cancer–killing virus in metastasized tumour tissue. These promising results could provide a foundation for systemic virotherapy for patients with cancer. SEE LETTER P. 99

EVANTHIA GALANIS

Clinical advances in cancer research are often slow to materialize, in part because the efficacy of a treatment has to be balanced against its potential toxicity to normal tissues. Infection of tumours with oncolytic (cancer-killing) viruses has been explored as a new type of treatment that is not cross-resistant with approved cancer therapies and, being target-specific, may have fewer toxic side effects. On page 99 of this issue, Breitbach *et al.*[1] describe a phase I clinical trial in which an intravenously delivered oncolytic poxvirus was capable of replicating selectively in metastasized tumours. This is a milestone in the development of an effective oncolytic agent for systemic administration.
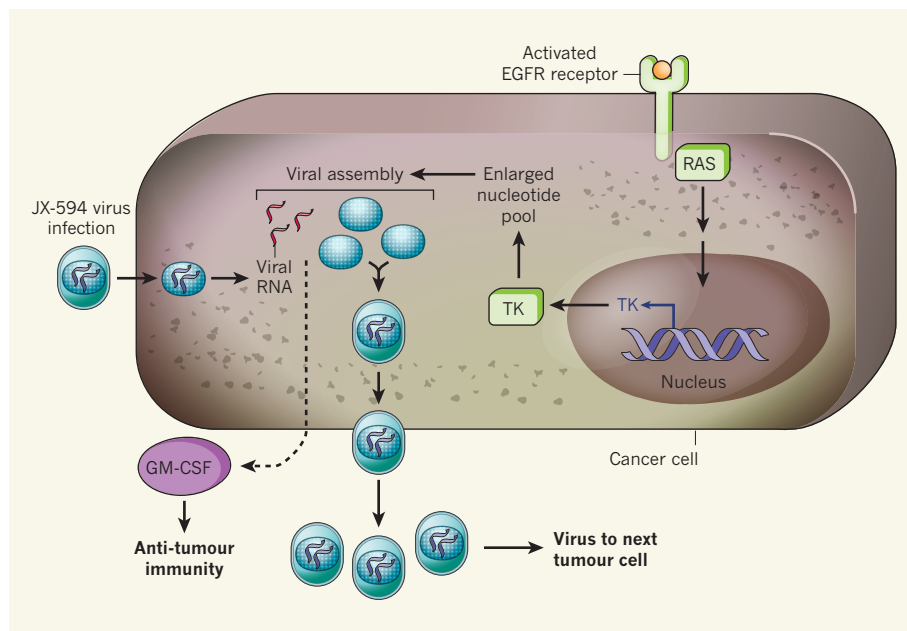
Oncolytic viruses became a focus of attention for cancer therapy following observations that natural viral infection or vaccination can lead to spontaneous regression of malignancies[2]. Unhindered by interferon-mediated antiviral defence, which is compromised in many tumours[3], these viruses specifically attack cancer cells by gaining entry through receptors that are overexpressed in these cells and/or by exploiting molecular pathways associated with malignant transformation for their replication[4,5]. As the virus starts to replicate at the tumour site, its destructive effect increases.

Strategies are being devised to make this process even more efficient by deploying genetically engineered oncolytic viruses that carry therapeutic or immunomodulatory transgenes.

In advanced cancer, systemic dissemination of solid tumours is linked with a poor prognosis. Before oncolytic viruses can be used to treat such metastases, they must be able to reach and replicate in metastatic sites following intravenous administration. But there are obstacles to be overcome, including the antiviral immune response, and the uptake and destruction of the virus by the endothelial reticulum system in the liver and spleen.

Breitbach *et al.*[1] take up the challenge using a genetically engineered oncolytic poxvirus known as JX-594. This is a smallpox-vaccine derivative of Wyeth-strain vaccinia virus carrying an inactivated thymidine kinase gene to increase tumour specificity, and expressing two transgenes: one encoding human granulocyte–macrophage colony-stimulating factor (GM-CSF) to stimulate anti-tumour immunity and the other β-galactosidase, a surrogate marker for detecting viral gene expression.

The authors tracked the virus in 23 cancer patients, all with advanced solid tumours that were resistant to other treatments. Patients were each given one dose of JX-594 at one of six different dosage levels by intravenous

injection; these were all well tolerated. The maximum feasible dose was 3×10^7 plaque-forming units (PFU) per kilogram of body weight (corresponding to a total dose of about 2×10^9 PFU). This dosage is in line with doses of other oncolytic viruses that can safely be given intravenously, including adenovirus, reovirus, paramyxovirus (Newcastle disease virus and measles) and Seneca Valley virus.

Breitbach *et al.* demonstrated such dose-dependent delivery of the virus (at 8–10 days after intravenous administration) to metastatic tumour deposits from a variety of tumour types, including leiomyosarcoma, mesothelioma, and lung, ovarian and colorectal cancers. In eight patients who had received 10^9 PFU or more per dose, delivery and replication were confirmed by quantitative polymerase chain reaction in five patients and by immunohistochemistry using a polyclonal anti-vaccinia antibody in six patients: granular cytoplasmic staining evident in tumour tissue was indicative of replicating virus (viral factories; Fig. 1).

Although JX-594 administration seemed to result in disease control in a dose-dependent way, with patients treated with the higher doses benefitting the most, viral infection and replication in metastatic deposits did not consistently affect clinical outcome. Some patients experienced clinical benefit — defined as disease stabilization for more than ten weeks — even when there was no evidence of viral replication in their tumour biopsies. By contrast, two out of six patients who were JX-594-positive by immunohistochemistry had progressive disease at first evaluation, even though replicating virus was detected in their metastatic tumours.

The explanation for these discrepancies may be down to several factors. For example, patients were allowed only one viral dose and treatment cycle: as with other cancer therapies, it is unlikely that a single round of treatment would be enough to stop tumour growth. Sampling variability in patients, whether

contain measurably different amounts of ^{15}N in their nitrogen pools. They found that, in the forest growing on nitrogen-rich sedimentary rock, the ^{15}N-content in both plants and soils matched that of the bedrock; this was not true for forests growing on the nitrogen-poor igneous rock, ruling out the possibility of significant nitrogen contribution from this rock.

Although the nitrogen-isotope measurements helped build the case for sedimentary bedrock as a nitrogen source for forests, they alone were not a smoking gun. To extend the findings beyond the carefully matched forest stands, the authors carried out a regional analysis of similar conifer forests in California. Sure enough, they found that the above-ground biomass of forests growing on nitrogen-rich sedimentary bedrock was almost 50% bigger by mass than that of forests on igneous bedrock,

after accounting for differing ages of tree stands.

The 'imprint' of nitrogen from bedrock on streams[5] and soils[6] has previously been reported for isolated sites in the same general region as the current study[3], and so Morford and colleagues' analysis makes the case for this as a regional pattern. But less than 2% of conifer-forest soils in that same region have a nitrogen capital as high as the sedimentary-bedrock forest that has been intensively studied by the authors (see Supplementary Information for ref. 3). This means that the high input of nitrogen from bedrock beneath that forest — which is equivalent to atmospheric nitrogen inputs — probably represents an upper estimate for the extent of this phenomenon. With 75% of Earth covered by sedimentary and related rock types[7], there is a real need to explore the phenomenon beyond this region to determine

what more common levels of bedrock-nitrogen inputs are for ecosystems elsewhere. ∎

Edward A. G. Schuur *is in the Department of Biology, University of Florida, Gainesville, Florida 32611-8526, USA.*
e-mail: tschuur@ufl.edu

1. Vitousek, P. M. & Howarth, R. W. *Biogeochemistry* **13,** 87–115 (1991).
2. Vitousek, P. M. *Nutrient Cycling and Limitation: Hawai'i as a Model System* (Princeton Univ. Press, 2004).
3. Morford, S. L., Houlton, B. Z. & Dahlgren, R. A. *Nature* **477,** 78–81 (2011).
4. Binkley, D., Son, Y. & Valentine, D. W. *Ecosystems* **3,** 321–331 (2000).
5. Holloway, J. M. *et al. Nature* **395,** 785–788 (1998).
6. Dahlgren, R. A. *Nature* **368,** 838–841 (1994).
7. Blatt, H. & Jones, R. L. *Geol. Soc. Am. Bull.* **86,** 1085–1088 (1975).

Tumour–fighting virus homes in

An early clinical trial demonstrates the delivery and replication of a cancer–killing virus in metastasized tumour tissue. These promising results could provide a foundation for systemic virotherapy for patients with cancer. SEE LETTER P.99

EVANTHIA GALANIS

Clinical advances in cancer research are often slow to materialize, in part because the efficacy of a treatment has to be balanced against its potential toxicity to normal tissues. Infection of tumours with oncolytic (cancer-killing) viruses has been explored as a new type of treatment that is not cross-resistant with approved cancer therapies and, being target-specific, may have fewer toxic side effects. On page 99 of this issue, Breitbach *et al.*[1] describe a phase I clinical trial in which an intravenously delivered oncolytic poxvirus was capable of replicating selectively in metastasized tumours. This is a milestone in the development of an effective oncolytic agent for systemic administration.

Oncolytic viruses became a focus of attention for cancer therapy following observations that natural viral infection or vaccination can lead to spontaneous regression of malignancies[2]. Unhindered by interferon-mediated antiviral defence, which is compromised in many tumours[3], these viruses specifically attack cancer cells by gaining entry through receptors that are overexpressed in these cells and/or by exploiting molecular pathways associated with malignant transformation for their replication[4,5]. As the virus starts to replicate at the tumour site, its destructive effect increases.

Strategies are being devised to make this process even more efficient by deploying genetically engineered oncolytic viruses that carry therapeutic or immunomodulatory transgenes.

In advanced cancer, systemic dissemination of solid tumours is linked with a poor prognosis. Before oncolytic viruses can be used to treat such metastases, they must be able to reach and replicate in metastatic sites following intravenous administration. But there are obstacles to be overcome, including the antiviral immune response, and the uptake and destruction of the virus by the endothelial reticulum system in the liver and spleen.

Breitbach *et al.*[1] take up the challenge using a genetically engineered oncolytic poxvirus known as JX-594. This is a smallpox-vaccine derivative of Wyeth-strain vaccinia virus carrying an inactivated thymidine kinase gene to increase tumour specificity, and expressing two transgenes: one encoding human granulocyte–macrophage colony-stimulating factor (GM-CSF) to stimulate anti-tumour immunity and the other β-galactosidase, a surrogate marker for detecting viral gene expression.

The authors tracked the virus in 23 cancer patients, all with advanced solid tumours that were resistant to other treatments. Patients were each given one dose of JX-594 at one of six different dosage levels by intravenous

injection; these were all well tolerated. The maximum feasible dose was 3×10^7 plaque-forming units (PFU) per kilogram of body weight (corresponding to a total dose of about 2×10^9 PFU). This dosage is in line with doses of other oncolytic viruses that can safely be given intravenously, including adenovirus, reovirus, paramyxovirus (Newcastle disease virus and measles) and Seneca Valley virus.

Breitbach *et al.* demonstrated such dose-dependent delivery of the virus (at 8–10 days after intravenous administration) to metastatic tumour deposits from a variety of tumour types, including leiomyosarcoma, mesothelioma, and lung, ovarian and colorectal cancers. In eight patients who had received 10^9 PFU or more per dose, delivery and replication were confirmed by quantitative polymerase chain reaction in five patients and by immunohistochemistry using a polyclonal anti-vaccinia antibody in six patients: granular cytoplasmic staining evident in tumour tissue was indicative of replicating virus (viral factories; Fig. 1).

Although JX-594 administration seemed to result in disease control in a dose-dependent way, with patients treated with the higher doses benefitting the most, viral infection and replication in metastatic deposits did not consistently affect clinical outcome. Some patients experienced clinical benefit — defined as disease stabilization for more than ten weeks — even when there was no evidence of viral replication in their tumour biopsies. By contrast, two out of six patients who were JX-594-positive by immunohistochemistry had progressive disease at first evaluation, even though replicating virus was detected in their metastatic tumours.

The explanation for these discrepancies may be down to several factors. For example, patients were allowed only one viral dose and treatment cycle: as with other cancer therapies, it is unlikely that a single round of treatment would be enough to stop tumour growth. Sampling variability in patients, whether

metastatic melanoma; the recently activated phase III trial testing addition of reovirus to paclitaxel/carboplatin chemotherapy in patients with recurrent head and neck cancer; and a randomized phase II trial comparing JX-594 with the best supportive care in patients with hepatocellular carcinoma for whom treatment with the drug sorafenib has failed.

In contrast to Asian countries, no virotherapy agent has so far been approved in the United States or Europe. The outcome of these trials may change this, generating additional valuable clinical tools for oncologists. ∎

Evanthia Galanis *is in the Mayo Clinic, Rochester, Minnesota 55905, USA. e-mail: galanis.evanthia@mayo.edu*

1. Breitbach, C. J. *et al. Nature* **477,** 99–102 (2011).
2. Liu, T. C., Galanis, E. & Kirn, D. *Nature Clin. Pract. Oncol.* **4,** 101–117 (2007).
3. Stojdl, D. F. *et al. Nature Med.* **6,** 821–825 (2000).
4. Katsafanas, G. C. & Moss, B. *J. Biol. Chem.* **279,** 52210–52217 (2004).
5. Norman, K. L., Hirasawa, K., Yang, A. D., Shields, M. A. & Lee, P. W. *Proc. Natl Acad. Sci. USA* **101,** 11099–11104 (2004).
6. Diop-Frimpong, B., Chauhan, V. P., Krane, S., Boucher, Y. & Jain, R. K. *Proc. Natl Acad. Sci. USA* **108,** 2909–2914 (2011).
7. Kottke, T. *et al. J. Clin. Invest.* **120,** 1551–1560 (2010).
8. Nguyen, T. L.-A. *et al. Proc. Natl Acad. Sci. USA* **105,** 14981–14986 (2008).
9. Wennier, S. T., Liu, J. & McFadden, G. *Curr. Pharm. Biotechnol.* (in the press).
10. Senzer, N. N. *et al. J. Clin. Oncol.* **27,** 5763–5771 (2009).

Figure 1 | Common oncogenic mutations in cancer cells encourage replication of the genetically engineered oncolytic JX-594 virus[1]. The virus takes advantage of a cancer cell's uncontrolled epidermal growth factor receptor (EGFR)–RAS signalling pathway. To replicate, this thymidine kinase (TK)-deficient virus relies on expression of TK by cancer cells. The newly assembled viruses then leave the cell to infect other tumour cells. These viruses also secrete GM-CSF, a factor that stimulates anti-tumour immunity. In normal cells, however, viral replication is blocked because this virus cannot efficiently exploit the cell's replication machinery.

positive or negative for JX-594, may also have confounded the results. Reassuringly, the normal tissue of patients in whom replication was detected was negative for replication by immunohistochemistry.

The limitations notwithstanding, these results convincingly demonstrate successful dose-dependent delivery and replication of an oncolytic virus in metastatic disease sites, following intravenous administration in patients with primary solid tumours. Although oncolytic viral replication in metastatic disease sites after systemic administration has been reported before, those studies are undermined by detectable replication only in isolated patients or by methodology unable to distinguish properly between input and progeny virus. Promising preclinical data, however, point to several strategies for enhancing systemic delivery of oncolytic viruses, including the use of cell carriers, cationic liposomes and polymers.

Large randomized trials to test oncolytic viruses in cancer treatment are ongoing or soon to be activated. These will investigate the potential synergistic cytotoxicity between oncolytic viruses and more conventional therapeutic approaches such as chemotherapy, small-molecule cell-cycle inhibitors, radiation therapy and anti-angiogenesis agents[6–9]. In addition, they will exploit induction of a systemic anti-tumour immune response in association with oncolytic tumour-cell death and expression of immunomodulatory transgenes[10].

Examples of such trials include the

soon-to-be-completed phase III trial of an attenuated strain of herpes simplex virus-1 that encodes GM-CSF in patients with

AGEING

Blood ties

The brain's ability to generate new neurons declines with age. This reduction is mediated by increased levels of an inflammatory factor in the blood of ageing mice and is associated with deficits in learning and memory. SEE LETTER P.90

RICHARD M. RANSOHOFF

On the face of it, the production of new neurons in the adult mammalian brain[1,2] sounds like a good thing. Interventions that reduce neurogenesis in adulthood can be associated with impaired brain function (in particular, with deficits in learning and memory), and the formation of neurons from neural stem cells declines with age. Understanding neurogenesis is therefore a major research goal, and neural stem cells are a tantalizing target for attempts to treat damaged brains by stimulating the production of neurons and other brain cells. On page 90 of this issue, Villeda and colleagues[3] report a crucial advance in this direction by identifying a blood-borne factor that affects neurogenesis and cognitive function in ageing mice.

With age, not only might the activity of neural stem cells (NSCs) deteriorate, but their immediate environment (the neurogenic niche) might also become compromised. The NSC niches lie near blood vessels, and factors that alter neurogenesis, such as exercise or systemic inflammation[4,5], might act by modifying blood cells or the abundance of signalling proteins in the blood plasma. Villeda *et al.* proposed, therefore, that agents present in the blood might affect neurogenesis.

To test this possibility, the authors[3] used a surgical procedure called parabiosis to connect the flank tissues of pairs of mice so that the animals developed a shared circulation. In mouse pairs of the same age (young–young or old–old), parabiosis alone did not affect neurogenesis. In the old–young pairs, however, the older animal showed enhanced neurogenesis, and in younger

Figure 1 | Common oncogenic mutations in cancer cells encourage replication of the genetically engineered oncolytic JX-594 virus[1]. The virus takes advantage of a cancer cell's uncontrolled epidermal growth factor receptor (EGFR)–RAS signalling pathway. To replicate, this thymidine kinase (TK)-deficient virus relies on expression of TK by cancer cells. The newly assembled viruses then leave the cell to infect other tumour cells. These viruses also secrete GM-CSF, a factor that stimulates anti-tumour immunity. In normal cells, however, viral replication is blocked because this virus cannot efficiently exploit the cell's replication machinery.

positive or negative for JX-594, may also have confounded the results. Reassuringly, the normal tissue of patients in whom replication was detected was negative for replication by immunohistochemistry.

The limitations notwithstanding, these results convincingly demonstrate successful dose-dependent delivery and replication of an oncolytic virus in metastatic disease sites, following intravenous administration in patients with primary solid tumours. Although oncolytic viral replication in metastatic disease sites after systemic administration has been reported before, those studies are undermined by detectable replication only in isolated patients or by methodology unable to distinguish properly between input and progeny virus. Promising preclinical data, however, point to several strategies for enhancing systemic delivery of oncolytic viruses, including the use of cell carriers, cationic liposomes and polymers.

Large randomized trials to test oncolytic viruses in cancer treatment are ongoing or soon to be activated. These will investigate the potential synergistic cytotoxicity between oncolytic viruses and more conventional therapeutic approaches such as chemotherapy, small-molecule cell-cycle inhibitors, radiation therapy and anti-angiogenesis agents[6–9]. In addition, they will exploit induction of a systemic anti-tumour immune response in association with oncolytic tumour-cell death and expression of immunomodulatory transgenes[10].

Examples of such trials include the

soon-to-be-completed phase III trial of an attenuated strain of herpes simplex virus-1 that encodes GM-CSF in patients with

metastatic melanoma; the recently activated phase III trial testing addition of reovirus to paclitaxel/carboplatin chemotherapy in patients with recurrent head and neck cancer; and a randomized phase II trial comparing JX-594 with the best supportive care in patients with hepatocellular carcinoma for whom treatment with the drug sorafenib has failed.

In contrast to Asian countries, no virotherapy agent has so far been approved in the United States or Europe. The outcome of these trials may change this, generating additional valuable clinical tools for oncologists. ■

Evanthia Galanis is in the Mayo Clinic, Rochester, Minnesota 55905, USA.
e-mail: galanis.evanthia@mayo.edu

1. Breitbach, C. J. *et al. Nature* **477,** 99–102 (2011).
2. Liu, T. C., Galanis, E. & Kirn, D. *Nature Clin. Pract. Oncol.* **4,** 101–117 (2007).
3. Stojdl, D. F. *et al. Nature Med.* **6,** 821–825 (2000).
4. Katsafanas, G. C. & Moss, B. *J. Biol. Chem.* **279,** 52210–52217 (2004).
5. Norman, K. L., Hirasawa, K., Yang, A. D., Shields, M. A. & Lee, P. W. *Proc. Natl Acad. Sci. USA* **101,** 11099–11104 (2004).
6. Diop-Frimpong, B., Chauhan, V. P., Krane, S., Boucher, Y. & Jain, R. K. *Proc. Natl Acad. Sci. USA* **108,** 2909–2914 (2011).
7. Kottke, T. *et al. J. Clin. Invest.* **120,** 1551–1560 (2010).
8. Nguyen, T. L.-A. *et al. Proc. Natl Acad. Sci. USA* **105,** 14981–14986 (2008).
9. Wennier, S. T., Liu, J. & McFadden, G. *Curr. Pharm. Biotechnol.* (in the press).
10. Senzer, N. N. *et al. J. Clin. Oncol.* **27,** 5763–5771 (2009).

AGEING

Blood ties

The brain's ability to generate new neurons declines with age. This reduction is mediated by increased levels of an inflammatory factor in the blood of ageing mice and is associated with deficits in learning and memory. SEE LETTER P.90

RICHARD M. RANSOHOFF

On the face of it, the production of new neurons in the adult mammalian brain[1,2] sounds like a good thing. Interventions that reduce neurogenesis in adulthood can be associated with impaired brain function (in particular, with deficits in learning and memory), and the formation of neurons from neural stem cells declines with age. Understanding neurogenesis is therefore a major research goal, and neural stem cells are a tantalizing target for attempts to treat damaged brains by stimulating the production of neurons and other brain cells. On page 90 of this issue, Villeda and colleagues[3] report a crucial advance in this direction by identifying a blood-borne factor that affects neurogenesis and cognitive function in ageing mice.

With age, not only might the activity of neural stem cells (NSCs) deteriorate, but their immediate environment (the neurogenic niche) might also become compromised. The NSC niches lie near blood vessels, and factors that alter neurogenesis, such as exercise or systemic inflammation[4,5], might act by modifying blood cells or the abundance of signalling proteins in the blood plasma. Villeda *et al.* proposed, therefore, that agents present in the blood might affect neurogenesis.

To test this possibility, the authors[3] used a surgical procedure called parabiosis to connect the flank tissues of pairs of mice so that the animals developed a shared circulation. In mouse pairs of the same age (young–young or old–old), parabiosis alone did not affect neurogenesis. In the old–young pairs, however, the older animal showed enhanced neurogenesis, and in younger

mice neurogenesis was reduced. Also, learning and memory were impaired in young mice after they received injections of plasma from old, but not young, mice.

What are the factors in the blood that differ between young and old mice and so affect neurogenesis? Villeda and co-workers ruled out direct effects of cells migrating from old to young mice in the parabiotic pairs, because cells from the older mice cannot enter the brain of young mice. They therefore compared a subset of plasma proteins — 66 to be exact — between young and old mice, evaluating young–old parabiotic pairs to see which proteins were more abundant in the older animals but reduced in the younger mice. They fixed on an unlikely culprit for altering neurogenesis, the chemokine protein CCL11.

Chemokines constitute a genetically and structurally coherent group of immune mediator molecules called cytokines. But although they were originally discovered through their ability to direct the migration of inflammatory white blood cells, chemokines are now recognized as being regulatory factors in the development of tissues as diverse as the central nervous system and the urogenital system[6]. One chemokine, CXCL12, and its receptors CXCR4 and CXCR7, have been accorded pride of place in neurogenic-niche physiology because of their well-known roles in the development and function of the central nervous system[7–9]. CCL11, by contrast, has been mainly linked to allergic conditions such as asthma. So, do the levels of this chemokine simply correlate with an anti-neurogenic environment, or could it be that CCL11 actively affects neurogenesis?

Villeda *et al.*[3] provide several lines of evidence suggesting the latter possibility. When the authors injected this chemokine systemically into young animals, neurogenesis was reduced. Moreover, antibodies that neutralize CCL11, when co-injected with the chemokine systemically or into the neurogenic niche itself, reversed this decline in neurogenesis. Furthermore, brain slices from mice given CCL11 injections showed reduced long-term potentiation (LTP), a neurophysiological correlate of learning.

Exactly how CCL11 affects neurogenesis and cognitive function remains unclear. *In vivo*, CCR3, the receptor through which CCL11 signals, has not been reproducibly identified on NSCs or on the neural progenitor cells that arise from them; so it is likely that an indirect mechanism is at play. Previous work suggested that CCL11 could affect neurogenesis through several pathways. For instance, its receptor could be present on microglia,[10] the brain cells that can produce cytokines and that, under some conditions, impair neurogenesis[11]. Also, exposure of NSC-containing mixed-cell cultures to CCL11 leads to decreased proliferation of NSCs[12]. However, the CCL11/CCR3 signals may not always be deleterious: mice lacking CCR3 show greater neuronal loss than those with the receptor when the peripheral segment of their facial nerve is severed[13].

It is possible that CCL11 modifies the action of another cytokine. Myeloid cells in the meninges membranes lining the brain are maintained in a state of restrained inflammation through the action of the regulatory cytokine interleukin-4 (ref. 14); this state of muted inflammation promotes learning and memory. CCL11 suppresses the ability of interleukin-4 to restrain the inflammatory functions of myeloid cells[15] and might thus reduce neurogenesis, causing memory and learning deficits.

But regardless of the mechanisms involved, the good news from this report[3] is that NSCs in the ageing brain do not undergo irreversible decline and can respond to a favourable environment, which includes the circulation. The precise link between NSCs, neurogenesis and blood cells is probably complex, involving more than just CCL11. For example, an earlier study[16] found that individual mice from a genetically heterogeneous stock show widely variable levels of neurogenesis, and that the degree of neurogenesis is strongly correlated with the ratios of two subsets of T cells in the blood — those expressing the CD4 or CD8 marker proteins — but is weakly correlated with a variety of behavioural tasks. What's more, altering the ratios of CD4- and CD8-bearing T cells modulates neurogenesis, rather than vice versa.

So it seems that there is a much more robust connection than previously suspected between the sites of neurogenesis in the adult brain and the systemic circulating pool of cells and proteins. Given the difficulty of manipulating neurogenic niches directly, this information is encouraging and should inspire increased activity in both joggers and neuroscientists. ∎

Richard M. Ransohoff *is in the Neuroinflammation Research Center, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio 44195, USA.*
e-mail: ransohr@ccf.org

1. Lazarini, F. & Lledo, P. M. *Trends Neurosci.* **34**, 20–30 (2011).
2. Aimone, J. B., Deng, W. & Gage, F. H. *Trends Cogn. Sci.* **14**, 325–337 (2010).
3. Villeda, S. A. *et al. Nature* **477**, 90–94 (2011).
4. Vallières, L., Campbell, I. L., Gage, F. H. & Sawchenko, P. E. *J. Neurosci.* **22**, 486–492 (2002).
5. Kronenberg, G. *et al. Neurobiol. Aging* **27**, 1505–1513 (2006).
6. Ransohoff, R. M. *Immunity* **31**, 711–721 (2009).
7. Dambly-Chaudière, C., Cubedo, N. & Ghysen, A. *BMC Dev. Biol.* **7**, 23 (2007).
8. Wang, Y. *et al. Neuron* **69**, 61–76 (2011).
9. Li, M. & Ransohoff, R. M. *Prog. Neurobiol.* **84**, 116–131 (2008).
10. He, J. *et al. Nature* **385**, 645–649 (1997).
11. Choi, S. H. *et al. Neuron* **59**, 568–580 (2008).
12. Krathwohl, M. D. & Kaiser, J. L. *Stem Cells* **22**, 109–118 (2004).
13. Wainwright, D. A. *et al. ASN Neuro* **1**, e00024 (2009).
14. Derecki, N. C. *et al. J. Exp. Med.* **207**, 1067–1080 (2010).
15. Stevenson, N. J. *et al. J. Leukoc. Biol.* **85**, 289–297 (2009).
16. Huang, G. J. *et al. PLoS Biol.* **8**, e1000561 (2010).

MATERIALS SCIENCE

Dry solution to a sticky problem

Sticking plasters revolutionized the protection of minor wounds, but they're not ideal for fragile skin. A material that mimics the adhesive properties of certain beetles' feet might provide a solution.

JEFFREY M. KARP & ROBERT LANGER

Adhesives that stick to skin for long periods of time, or over multiple cycles of use, are vital for medical applications. Such materials have to conform to stringent standards — for example, they must maintain robust adhesion during repeated application and removal without irritating the skin, and be non-toxic. Writing in *Advanced Materials*, Kwak *et al.*[1] report an exciting advance towards achieving these standards: an adhesive tape that uses micrometre-scale pillars on its surface to stick to skin. This innovation bypasses the need for a glue-coated surface, as is commonly used in conventional skin adhesives.
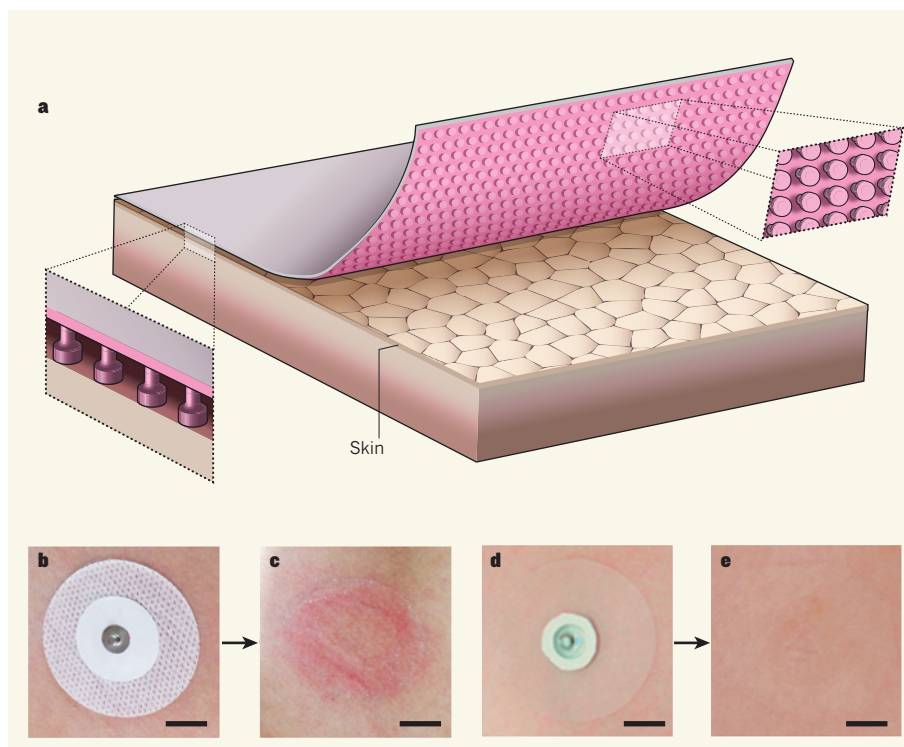
Skin adhesives are currently used billions of times a year — for example, in over-the-counter sticking plasters for the treatment of minor skin wounds, in transdermal patches for controlled drug delivery[2,3], and in tapes for affixing tubes or sensors to the skin in hospitals. Despite the remarkable success of these materials, a remaining challenge is to find adhesives suitable for use on the delicate skin of newborn infants and the elderly. Aged skin is particularly fragile, making it more susceptible to inflammation and damage[4]. Given that the number of people aged over 60 will double during the next two to three decades[5], the need for skin adhesives for the elderly is becoming increasingly pressing.

mice neurogenesis was reduced. Also, learning and memory were impaired in young mice after they received injections of plasma from old, but not young, mice.

What are the factors in the blood that differ between young and old mice and so affect neurogenesis? Villeda and co-workers ruled out direct effects of cells migrating from old to young mice in the parabiotic pairs, because cells from the older mice cannot enter the brain of young mice. They therefore compared a subset of plasma proteins — 66 to be exact — between young and old mice, evaluating young–old parabiotic pairs to see which proteins were more abundant in the older animals but reduced in the younger mice. They fixed on an unlikely culprit for altering neurogenesis, the chemokine protein CCL11.

Chemokines constitute a genetically and structurally coherent group of immune mediator molecules called cytokines. But although they were originally discovered through their ability to direct the migration of inflammatory white blood cells, chemokines are now recognized as being regulatory factors in the development of tissues as diverse as the central nervous system and the urogenital system[6]. One chemokine, CXCL12, and its receptors CXCR4 and CXCR7, have been accorded pride of place in neurogenic-niche physiology because of their well-known roles in the development and function of the central nervous system[7–9]. CCL11, by contrast, has been mainly linked to allergic conditions such as asthma. So, do the levels of this chemokine simply correlate with an anti-neurogenic environment, or could it be that CCL11 actively affects neurogenesis?

Villeda *et al.*[3] provide several lines of evidence suggesting the latter possibility. When the authors injected this chemokine systemically into young animals, neurogenesis was reduced. Moreover, antibodies that neutralize CCL11, when co-injected with the chemokine systemically or into the neurogenic niche itself, reversed this decline in neurogenesis. Furthermore, brain slices from mice given CCL11 injections showed reduced long-term potentiation (LTP), a neurophysiological correlate of learning.

Exactly how CCL11 affects neurogenesis and cognitive function remains unclear. *In vivo*, CCR3, the receptor through which CCL11 signals, has not been reproducibly identified on NSCs or on the neural progenitor cells that arise from them; so it is likely that an indirect mechanism is at play. Previous work suggested that CCL11 could affect neurogenesis through several pathways. For instance, its receptor could be present on microglia,[10] the brain cells that can produce cytokines and that, under some conditions, impair neurogenesis[11]. Also, exposure of NSC-containing mixed-cell cultures to CCL11 leads to decreased proliferation of NSCs[12]. However, the CCL11/CCR3 signals may not always be deleterious: mice

lacking CCR3 show greater neuronal loss than those with the receptor when the peripheral segment of their facial nerve is severed[13].

It is possible that CCL11 modifies the action of another cytokine. Myeloid cells in the meninges membranes lining the brain are maintained in a state of restrained inflammation through the action of the regulatory cytokine interleukin-4 (ref. 14); this state of muted inflammation promotes learning and memory. CCL11 suppresses the ability of interleukin-4 to restrain the inflammatory functions of myeloid cells[15] and might thus reduce neurogenesis, causing memory and learning deficits.

But regardless of the mechanisms involved, the good news from this report[3] is that NSCs in the ageing brain do not undergo irreversible decline and can respond to a favourable environment, which includes the circulation. The precise link between NSCs, neurogenesis and blood cells is probably complex, involving more than just CCL11. For example, an earlier study[16] found that individual mice from a genetically heterogeneous stock show widely variable levels of neurogenesis, and that the degree of neurogenesis is strongly correlated with the ratios of two subsets of T cells in the blood — those expressing the CD4 or CD8 marker proteins — but is weakly correlated with a variety of behavioural tasks. What's more, altering the ratios of CD4- and CD8-bearing T cells modulates neurogenesis, rather than vice versa.

So it seems that there is a much more robust connection than previously suspected between the sites of neurogenesis in the adult brain and the systemic circulating pool of cells and proteins. Given the difficulty of manipulating neurogenic niches directly, this information is encouraging and should inspire increased activity in both joggers and neuroscientists. ∎

Richard M. Ransohoff *is in the Neuroinflammation Research Center, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio 44195, USA.*
e-mail: ransohr@ccf.org

1. Lazarini, F. & Lledo, P. M. *Trends Neurosci.* **34,** 20–30 (2011).
2. Aimone, J. B., Deng, W. & Gage, F. H. *Trends Cogn. Sci.* **14,** 325–337 (2010).
3. Villeda, S. A. *et al. Nature* **477,** 90–94 (2011).
4. Vallières, L., Campbell, I. L., Gage, F. H. & Sawchenko, P. E. *J. Neurosci.* **22,** 486–492 (2002).
5. Kronenberg, G. *et al. Neurobiol. Aging* **27,** 1505–1513 (2006).
6. Ransohoff, R. M. *Immunity* **31,** 711–721 (2009).
7. Dambly-Chaudière, C., Cubedo, N. & Ghysen, A. *BMC Dev. Biol.* **7,** 23 (2007).
8. Wang, Y. *et al. Neuron* **69,** 61–76 (2011).
9. Li, M. & Ransohoff, R. M. *Prog. Neurobiol.* **84,** 116–131 (2008).
10. He, J. *et al. Nature* **385,** 645–649 (1997).
11. Choi, S. H. *et al. Neuron* **59,** 568–580 (2008).
12. Krathwohl, M. D. & Kaiser, J. L. *Stem Cells* **22,** 109–118 (2004).
13. Wainwright, D. A. *et al. ASN Neuro* **1,** e00024 (2009).
14. Derecki, N. C. *et al. J. Exp. Med.* **207,** 1067–1080 (2010).
15. Stevenson, N. J. *et al. J. Leukoc. Biol.* **85,** 289–297 (2009).
16. Huang, G. J. *et al. PLoS Biol.* **8,** e1000561 (2010).

MATERIALS SCIENCE

Dry solution to a sticky problem

Sticking plasters revolutionized the protection of minor wounds, but they're not ideal for fragile skin. A material that mimics the adhesive properties of certain beetles' feet might provide a solution.

JEFFREY M. KARP & ROBERT LANGER

Adhesives that stick to skin for long periods of time, or over multiple cycles of use, are vital for medical applications. Such materials have to conform to stringent standards — for example, they must maintain robust adhesion during repeated application and removal without irritating the skin, and be non-toxic. Writing in *Advanced Materials*, Kwak *et al.*[1] report an exciting advance towards achieving these standards: an adhesive tape that uses micrometre-scale pillars on its surface to stick to skin. This innovation bypasses the need for a glue-coated surface, as is commonly used in conventional skin adhesives.

Skin adhesives are currently used billions of times a year — for example, in over-the-counter sticking plasters for the treatment of minor skin wounds, in transdermal patches for controlled drug delivery[2,3], and in tapes for affixing tubes or sensors to the skin in hospitals. Despite the remarkable success of these materials, a remaining challenge is to find adhesives suitable for use on the delicate skin of newborn infants and the elderly. Aged skin is particularly fragile, making it more susceptible to inflammation and damage[4]. Given that the number of people aged over 60 will double during the next two to three decades[5], the need for skin adhesives for the elderly is becoming increasingly pressing.

Pressure-sensitive surgical tapes first appeared in 1845, when the surgeon Horace Day applied rubber adhesive to strips of fabric[6]. For several years thereafter, minor cuts were treated with separate gauze and adhesive tape, but custom tailoring of the materials was required for domestic use. The first integrated skin-adhesive device — the Band-Aid — was invented in 1920 by Earle Dickson[7], an employee at the company Johnson & Johnson. Dickson noticed that gauze and adhesive tape did not remain attached to his wife's fingers, which she frequently injured in the kitchen. He therefore placed gauze in the centre of a strip of tape and covered the adhesive and gauze with a layer of crinoline to maintain its tack and sterility. Johnson & Johnson began mass-producing these sticking plasters shortly thereafter, and today it is estimated that more than 100 billion of Dickson's Band-Aids have been made[7].

The adhesives currently used for sticking plasters are polymeric, pressure-sensitive adhesives based on acrylic compounds[8]. Although effective, acrylic adhesives can leave behind sticky residues, and they lose their grip after repeated use. To bypass the need for these glues, researchers have focused on adhesion mechanisms used by animals such as beetles and geckos, whose feet stick to walls without any glue. The mechanism of gecko-foot adhesion was elucidated[9] in 2000, nearly two millennia after Aristotle first reported the phenomenon: each gecko foot contains up to 500,000 hairs, each tipped with hundreds of projections known as spatulae. Similarly, the feet of beetles in the Chrysomelidae family are covered with tiny mushroom-shaped structures that help them cling to surfaces.

Gecko spatulae are roughly hundreds of nanometres in length, whereas the mushroom-shaped structures of Chrysomelidae beetles' feet are on the micrometre scale. It is possible to mimic these adhesive structures using nanometre- or micrometre-scale engineering to modify the surfaces of materials. Synthetic gecko-inspired adhesives have been made, but it has been difficult to optimize their properties for successful adhesion to wet tissues (such as those found inside the body). To solve this problem, we have previously used a hybrid approach, whereby a rubbery polymeric substrate with the surface nanotopography of gecko feet was coated with a thin layer of tissue-reactive glue[10]. The resulting material maximized adhesion to wet tissue while minimizing tissue inflammation.

Kwak et al.[1] have focused on achieving adhesion to dry skin in the absence of glue. They patterned the surface of a rubbery, non-toxic substrate with micrometre-scale, mushroom-shaped projections (Fig. 1) — a topology reported to be ideal for maximizing adhesion[11] — varying the dimensions of the projections until they achieved optimal adhesion to human skin in a direction perpendicular to its surface. Remarkably, the substrate

maintained good adhesion through up to 30 cycles of attachment and removal, without causing significant damage to skin.

To demonstrate the functional utility of their adhesive, the authors integrated it into a wearable diagnostic device that monitors the heart using electrocardiography. When attached to a patient's chest, the device recorded several vital signals from the heart in real time over a period of two days. For commercial applications, however, Kwak and colleagues' material will probably require higher levels of adhesion — the reported system[1] achieved about 43% of the adhesion of a moderately sticky acrylic.

The authors' work is part of a growing body of research aimed at finding new materials that form interfaces with tissue. For example, another recent paper[12] describes single-use ultrathin membranes that adhere to skin using only van der Waals interactions, and which incorporate electronic components that can be used to perform electrophysiological recordings. For long-term applications, these technologies should be tested both in the presence of humidity or perspiration and to see how they cope with the shedding of dead cells. New approaches may be required to address such issues, perhaps involving surface-responsive materials[13]. Nevertheless, there is every hope that innovations such as that of Kwak et al.[1] will one day bring new technologies to the bedside. ■

Figure 1 | Glueless. a, Kwak et al.[1] have made a polymer material that has a surface covered in micrometre-scale, mushroom-shaped projections (upper inset). The projections mimic those found on certain beetles' feet, and allow the substrate to adhere to human skin without the use of glue (lower inset). b, c, Commonly used sticking plasters (b) use an acrylic adhesive to stick to skin, but can leave behind a sticky residue and cause redness (c). d, e, A patch made from Kwak and colleagues' material (d) reduces these effects (e). Scale bars, 1 cm. (Graphic adapted from ref. 1. Photos reproduced from ref. 1.)

Jeffrey M. Karp is at the Center for Regenerative Therapeutics, and the Department of Medicine, Brigham and Women's Hospital, Cambridge, Massachusetts 02139, USA. Robert Langer is at the David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mails: jkarp@rics.bwh.harvard.edu; rlanger@mit.edu

1. Kwak, M. K., Jeong, H.-E. & Suh, K. Y. Adv. Mater. http://dx.doi.org/10.1002/adma.201101694 (2011).
2. Prausnitz, M. R., Mitragotri, S. & Langer, R. Nature Rev. Drug Discov. 3, 115–124 (2004).
3. Prausnitz, M. R. & Langer, R. Nature Biotechnol. 26, 1261–1268 (2008).
4. Bianchi, J. & Cameron, J. Br. J. Community Nurs. 13, Issue 3 suppl., S26–S32 (2008).
5. www.un.org/esa/population/publications/worldageing19502050/pdf/80chapterii.pdf
6. Smith, M. A., Jones, N. M. M., Page, S. L. & Dirda, M. P. J. Am. Inst. Conserv. 23, 101–113 (1984).
7. http://web.mit.edu/invent/iow/dickson.html
8. Auchter, G., Aydin, O., Zetti, A. & Satas, D. in Handbook of Pressure Sensitive Adhesive Technology 3rd edn (ed. Satas, D.) 444–514 (Satas & Associates, 1999).
9. Autumn, K. et al. Nature 405, 681–685 (2000).
10. Mahdavi, A. et al. Proc. Natl Acad. Sci. USA 105, 2307–2312 (2008).
11. Carbone, G., Pierro, E. & Gorb, S. N. Soft Matter 7, 5545–5552 (2011).
12. Kim, D.-H. et al. Science 333, 838–843 (2011).
13. Chan, E. P., Karp, J. M. & Langer, R. S. J. Polym. Sci. B 49, 40–44 (2011).

REVIEW

Synthesis, assembly and applications of semiconductor nanomembranes

J. A. Rogers[1], M. G. Lagally[2] & R. G. Nuzzo[3]

Research in electronic nanomaterials, historically dominated by studies of nanocrystals/fullerenes and nanowires/ nanotubes, now incorporates a growing focus on sheets with nanoscale thicknesses, referred to as nanomembranes. Such materials have practical appeal because their two-dimensional geometries facilitate integration into devices, with realistic pathways to manufacturing. Recent advances in synthesis provide access to nanomembranes with extraordinary properties in a variety of configurations, some of which exploit quantum and other size-dependent effects. This progress, together with emerging methods for deterministic assembly, leads to compelling opportunities for research, from basic studies of two-dimensional physics to the development of applications of heterogeneous electronics.

Semiconductor nanomembranes (NMs) are monocrystalline structures with thicknesses of less than a few hundred nanometres and with minimum lateral dimensions at least two orders of magnitude larger than the thickness. They differ from thin films in that they exist as free-standing, isolated forms at some critical stage in their growth or processing, or in their final, device-integrated forms. Because NMs offer many features that cannot be reproduced in other material formats, they are of central importance to a rapidly expanding frontier in nanoscience and technology. The origins of work on NMs can be traced back nearly thirty years to exploratory research on cadmium-based nanocrystals[1] and spherical fullerenes[2]. Studies of these and other 'zero-dimensional' materials evolved to include nanowires and carbon nanotubes[3], partly because it is comparatively easy (although still difficult) to form electrical contacts to such 'one-dimensional' structures. Although diverse types of semiconductor device are possible with individual wires/tubes, their practical application in high-yield, scalable systems faces formidable engineering challenges in assembly and other aspects of manufacturing. Materials in NM formats avoid these limitations, because their two-dimensional (2D) geometries are directly compatible with established device designs and processing approaches from the semiconductor industry, building naturally on decades of research in thin-film growth, patterning and processing. NMs also have finite-size and quantum characteristics in their electronic, phononic and optical properties, and have unique mechanical features, with effects related to shape distortions and folding of sheets, not found in zero- and one-dimensional materials. NMs can be made uniformly and repeatably (in size, shape, surface orientation, thickness and surface roughness) by 'top-down" methods used in semiconductor device manufacture.

Many advanced materials have begun to be studied in this format. These include a wide variety of inorganic materials, from those as common as silicon to esoteric layered compounds[4–6], as well as a rapidly growing range of forms of conjugated carbon[7–9], not limited to graphene. Sophisticated methods are becoming available for manipulating NMs with thicknesses of as little as a fraction of a nanometre and with lateral dimensions of up to many square centimetres, at high throughputs and yields[10–12]. NMs can be distributed over large areas, folded into various shapes[13–16] and wrapped onto curvilinear surfaces[17]. Advanced electronic and optoelectronic devices have been reported, each with a unique combination of operating speed[18], heterogeneous layout[19], flexible design[20,21], three-dimensional (3D) form[17,19] and other features that would be difficult or impossible to achieve with existing bulk materials technologies or with zero- or one-dimensional nanomaterials. These advances motivate the present Review of approaches to synthesis, assembly, and device integration for inorganic and organic NMs, excluding graphene, with an emphasis on challenges and future opportunities.
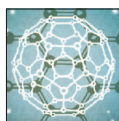
Inorganic nanomembranes

Single-crystalline inorganic semiconductor NMs with thicknesses that match length scales of important physical processes (a few hundred nanometres or less) offer opportunities in basic and applied research, as well as in technology, as suggested by recent demonstrations of practical devices that offer operational features unavailable with bulk materials. Figure 1 illustrates representative examples of NM properties in mechanics, electronics, thermoelectrics and photonics. In the first instance (Fig. 1a), the extremely small thicknesses of NMs (down to ~2 nm for silicon) lead to flexural rigidities that can be more than fifteen orders of magnitude smaller than those of bulk wafers (~200 µm) of the same materials[22]. The resulting values are so small, in fact, that they qualitatively change the nature of the material to allow otherwise impossible, non-planar geometries and multi-layer integration options. The latter capabilities arise from the combined effects of low rigidities and energy release rates for thermally driven delamination that decrease linearly with thickness[23]. As a consequence, NMs conform and bond robustly to nearly any surface, thereby enabling them to be stacked onto one another or onto foreign hosts to yield unusual, heterogeneous systems that cannot be achieved with wafer-bonding technologies or epitaxy. Such stacking can lead to unusual electronic, electromechanical, thermoelectric, optoelectronic, optomechanical and photonic behaviour.

Sufficient thinness yields the 2D physics of quantum confinement, even in simple, single-layer NMs, with important implications for electronic transport. Figure 1b shows, as an example, the splitting of the conduction-band-minimum valley of silicon into subbands as a function of the thickness of the NM, for two orientations[24]. The right-hand plot shows how NM roughness affects the 2D density of states for these quantum-confined NMs[24]. A related phenomenon is the extremely high sensitivity

a Mechanics

b Electronics

c Thermoelectrics

Photonics

d
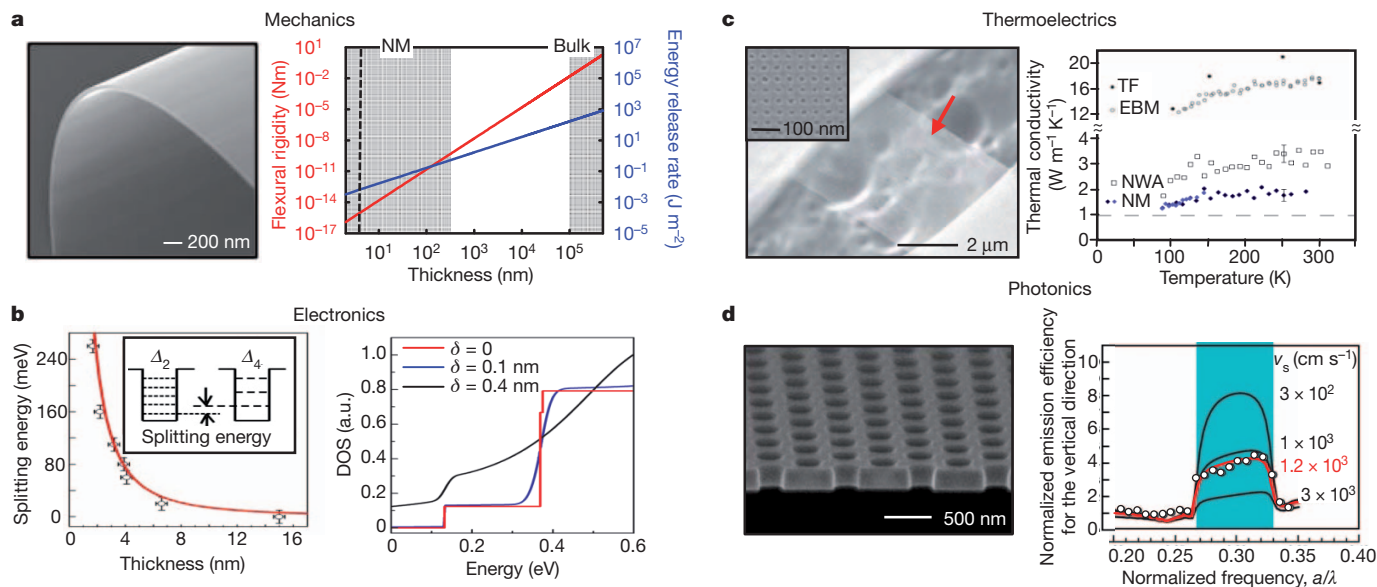
Figure 1 | Unique physical properties in NMs. a, NMs have exceptionally high degrees of bendability, as illustrated in the scanning electron microscope (SEM) image. The flexural rigidity of a 2-nm-thick, silicon NM is $\sim 10^{15}$ times smaller than that of its bulk wafer counterpart (200 μm thick), as illustrated in the red curve of the graph (dashed line at 2 nm). Related mechanics allows bonding of NMs to nearly any surface. Here energy release rates associated with opening of interfaces between NMs and supporting substrates decrease linearly with thickness. The blue line represents calculations for silicon NMs bonded to sheets of polyimide at room temperature, and then heated to 300 °C. b, Electronic confinement effects in silicon NMs lead to splitting of the conduction band valleys (Δ) for the (001) orientation (left) with representative 1-s.d. error bars. Here the surface roughness (δ) strongly affects the 2D density of states[24] (DOS; right). a.u., arbitrary units. c, Phonon confinement in NMs offers opportunities for manipulating heat flow, to optimize figures of merit in thermoelectrics. The image shows a suspended silicon NM (22 nm thick; red

arrow) perforated with arrays of nanoholes (diameter, ~ 10–15 nm; period, ~ 35 nm) that scatter phonons, thereby frustrating thermal transport[26]. The data compare such structures (NM) with arrays of nanowires (NWA; 28 nm wide, 20 nm thick), coarsely patterned NMs (EBM; square mesh with period of 385 nm, 22 nm thick) and uniform NMs (TF; 25 nm thick), with representative 1-s.d. error bars[26]. d, Photon confinement in NMs allows for low-threshold lasers. The SEM image shows a photonic crystal that consists of an array of nanoholes (period, ~ 500 nm) in a GaInAsP NM (245 nm thick), which is designed to suppress rates of spontaneous emission and, simultaneously, to direct light into vertical modes[28]. The graph shows measurements[28] (symbols) of emission efficiency, normalized to the case without nanoholes, as a function of the ratio of the period of the array (a) to the emission wavelength (λ). The results indicate enhancements for a range of a/λ values. Calculations (solid lines) with various surface-recombination velocities (v_s) capture the trends[28]. The blue region corresponds to the location of the photonic bandgap.

of charge transport to surface chemical condition, which also modifies the local band structure[25]. Other examples of dimensional effects in electronic properties are discussed below, in sections on synthesis and applications. The small thicknesses of NMs also strongly influence the behaviour of phonons and photons. Phonons can be used to increase key figures of merit (for example ZT) in thermoelectrics. NMs with arrays of nanoholes (Fig. 1c) that have lateral dimensions less than the mean free path of thermal phonons (~ 300 nm) and with thicknesses of about this value or less, produce strong backscattering effects that frustrate thermal transport, without reducing the electrical sheet resistance, owing to the comparatively shorter mean free paths of electrons and holes[26,27] (1–10 nm at high doping levels). Certain measurements of NMs with holes suggest thermal conductivities ~ 80 times smaller than values in bulk silicon, and enhancements in ZT of a factor of ~ 50 relative to NMs without holes[27]. Conceptually related effects of confinement can be used to advantage in optoelectronics. Some of the earliest examples involve lasers using NM-based photonic crystals that inhibit spontaneous emission by 94% and guide preferential emission into vertical modes, improving output efficiency by a factor of almost five[28,29] (Fig. 1d). The thicknesses of the NMs in such systems are typically a fraction of the emission wavelength to guarantee single-mode behaviour. Measurements and simulations shown in Fig. 1d capture these effects[28] and also underscore the critical role of surface recombination in NMs, as discussed below in the context of related applications.

Challenges common to all research efforts in NMs lie in the development of improved methods to synthesize NMs with precise dimensions and materials quality; to engineer structural features or distributions of strain in NMs to yield unusual charge, photon or phonon transport characteristics; to stack and bond NMs, with an emphasis on the control of surface/interfacial properties and on charge transport across bonded-NM

interfaces; to deform NMs into unusual shapes for non-planar components, such as cylindrical microcavity lasers and stretchable, bio-integrated electronics; and to produce large numbers of NMs with precise geometries, efficiently and cost effectively, and assemble them in desired configurations at high throughputs. The following sections highlight exemplary recent advances and areas of future opportunity.

Synthesis

In a route similar to that used to create graphene from bulk pieces of graphite, inorganic NMs can be (and have been for many years) formed by chemical or mechanical exfoliation from solids that have naturally layered structures[4–6], such as the semiconductors MoS_2 (refs 4, 30, 31), $Sr_2Nb_3O_{10}$ (ref. 5), GeS (ref. 32) and GeSe (ref. 32). Recent work indicates that confinement effects in single-layer NMs of MoS_2 (Fig. 2a) lead to direct bandgaps, unlike the indirect gaps of bulk samples[31]. Transistors that incorporate these NMs show field-effect mobilities greater than those that use graphene structured in nanoribbon geometries required to create a bandgap for efficient switching behaviour[31]. As a result, exfoliated NMs are of interest as ultrathin alternatives to graphene for active materials in next-generation electronics, where interband tunnelling might be used to improve performance in low-power devices further.

Although exfoliation yields large numbers of NMs from certain classes of semiconductors such as MoS_2, additional methods are needed to control their dimensions and shapes, and to manipulate them for integration into systems. A synthetic strategy that addresses these requirements and expands the materials options involves the release of NMs from bulk semiconductors that are not naturally layered, by use of specialized anisotropic etching procedures. For example, defining trenches on the surfaces of silicon wafers with {111} orientations, and
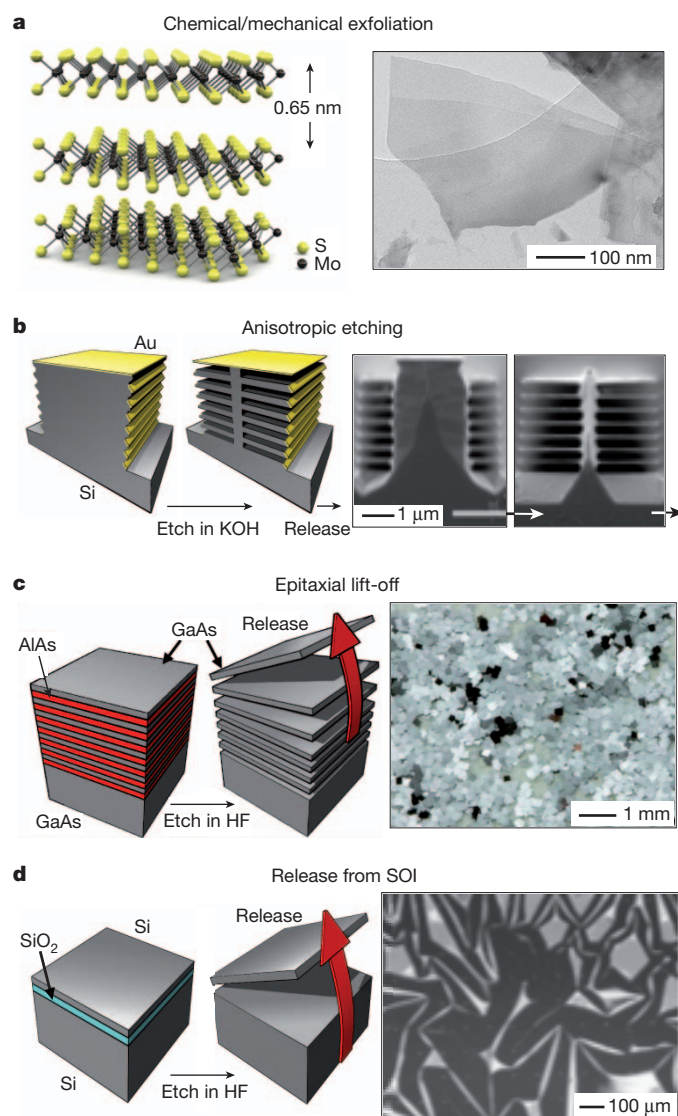
Figure 2 | Representative routes for synthesizing inorganic monocrystalline semiconductor NMs. a, Atomic structure of MoS_2, showing its layered configuration[31]. Chemical or mechanical exfoliation of this material yields single-layer NMs (0.65 nm thick), as shown in the transmission electron microscope image on the right[30]. **b**, Process for generating multilayer stacks of silicon NMs from a bulk wafer by anisotropic etching. Patterned features of etch resist (gold) on the structured sidewalls of vertically etched trenches allow access of an anisotropic wet chemical etchant only to certain regions of the silicon. Etching releases silicon NMs (~100 nm thick), as shown at two intermediate times in the cross-sectional SEM images on the right[35]. **c**, Epitaxial multilayer assembly of GaAs and aluminium arsenide (AlAs) grown on a GaAs wafer[36]. Etching vertically through the thickness of the stack and then immersing the structure in hydrofluoric acid leads to the selective removal of the AlAs layers. Complete undercut etching releases large numbers of GaAs NMs. The SEM image shows a collection of GaAs membranes formed using this process[36]. **d**, Release of a silicon NM from a SOI wafer. Etching vertically through the top silicon layer exposes the underlying SiO_2 layer, allowing its removal by etching in hydrofluoric acid. The optical image on the right shows a wrinkled, but completely single-crystal, silicon NM (~50 nm thick) formed in this manner that can then be transferred to a new host, where it will flatten and bond[10].

then patterning their side walls with etch resists provides a starting point for the anisotropic removal of material along the ⟨110⟩ directions using solutions of potassium hydroxide[33,34]. The process (Fig. 2b) releases stacks, and, with minor modifications, essentially bulk-like numbers of NMs[35]. An appealing aspect of this synthesis is that lithography defines the lateral dimensions of the NMs and their spatial positions across the wafer, thereby rendering them compatible with methods for

integration described below. Cycles of thermal oxidation and etching can reduce the thicknesses and passivate the interfaces.

Methods that offer atomically smooth surfaces and enhanced dimensional control over large areas begin with epitaxial growth to form releasable multilayer assemblies. For example, stacks of gallium arsenide (GaAs) films separated by aluminium gallium arsenide (AlGaAs) yield large numbers of GaAs NMs on selective removal of the AlGaAs with hydrofluoric acid[36]. This process provides significant cost and throughput advantages over related approaches that release only single layers[37–39], owing to more efficient use of the systems for epitaxial growth and the supporting substrates[36]. Fig. 2c shows an illustration of the process and an image of a deposit of released GaAs NMs formed by casting from a solution suspension[36]. Similar multilayers can yield, from a single stack, NMs for multifunctional integrated systems of radio-frequency electronics, photodetectors, light emitters and solar cells[36,40]. Epitaxial layers for passivation can be grown directly, to form an integral part of the NM. Multimaterial structures, such as those of GaAs with self-assembled quantum dots of indium arsenide (InAs) as embedded light emitters, are also possible, for applications in optoelectronics[41]. The thin geometries of NMs can be important in these types of epitaxial synthesis because they avoid the dislocations that occur in films thicker than the critical thickness for defect formation in strained layers[42]. Additionally, greater materials possibilities follow from the use of NMs as growth substrates, where modified lattice constants or strain symmetries, not achievable in bulk materials, can be exploited, as described below.

A complementary scheme that avoids the demands of epitaxy entirely uses layered materials formed by wafer bonding followed by polishing or controlled fracture. The most common example starts with a thin layer of silicon on silicon dioxide (SiO_2) supported by a silicon substrate, known as silicon on insulator[13,43–45] (SOI). Etching the buried oxide with hydrofluoric acid releases the top silicon layer as a NM (Fig. 2d). Commercially available SOI can be used to make silicon NMs with thicknesses down to 20 nm. Oxidation and etching can reduce the thickness to <2 nm, with uniformity greater than 0.3 nm (ref. 25). Other examples of SOI-like structures include group-IV analogues such as germanium on insulator, strained silicon on insulator and silicon–germanium (SiGe) on insulator, as well as III–V semiconductors and many other combinations[46–48].

Strain engineering and 3D nanoarchitectures

An intriguing aspect of the synthesis of NMs is the ability to modify structures by lithographic processing (Fig. 1c, d) or by introducing spatial distributions of strain. The latter offers great promise for the creation and investigation of new physical properties. The distinctive mechanical properties of NMs (see discussion above and Fig. 1a), allow this strain engineering. Strain changes the lattice constant, thereby creating new properties relative to the unstrained, but chemically identical, material. The ability to alter the strain, in magnitude, direction, spatial extent, periodicity, symmetry and/or nature, allows tuning of the intrinsic properties to such a degree that many are significantly modified, including band structure, charge carrier mobility, atomic transport, atomic defect structure, the self-assembly of quantum dots, piezoresistivity, and more complex phenomena such as electro-optical effects[49–51].

Lattice strain can be introduced by heteroepitaxial growth of materials with different lattice constants[13,44,52]. For example, in a trilayer of Si/SiGe/Si grown on SOI, the SiGe layer is compressively strained and the silicon layers are unstrained. When this trilayer is released, the SiGe layer shares its strain elastically with the silicon layers, with strain magnitudes of up to 1%. Strains in this range can cause significant changes in the band structure of silicon[50], such as to improve the performance of transistors on flexible polymers[53], and, if applied locally and periodically, to form strain-induced, single-element electronic heterojunction superlattices[49,54]. This elastic strain sharing, with appropriate processing, can be used to produce special defect-free semiconductors that cannot be realized in other ways[55], and, by taking advantage of differing crystal symmetries in the components of strained multilayer composites, to make entirely new

materials with crystal symmetries that do not exist in bulk form and cannot be created by heteroepitaxy alone[56].

Strain can also be introduced mechanically. Again, the distinctive mechanical properties of NMs are decisive in achieving novel properties. For example, it has recently become possible, because the conduction-band valleys shift by different amounts with strain, to stretch germanium NMs biaxially sufficiently to change germanium from an indirect-bandgap semiconductor to a direct-bandgap semiconductor[57]. This transformation allows use of germanium in light sources, thereby realizing the vision of group-IV-semiconductor integrated electronics and optics.

A related consequence of strain control is that the geometries of NMs can be engineered to yield 3D shapes, allowing device configurations and properties that would be impossible to achieve with bulk materials. Possibilities include tubes that can provide active growth platforms for cells[58], cylindrical microcavities that can serve as fluidic channels and optical sensors[59], and buckled structures that can respond elastically to large strains and can be used in stretchable electronics[45,60,61]. With appropriate engineering of strains in systems in which there is a strain gradient perpendicular to the layers, tubes, spirals, rings and other 3D nanoarchitectures can be achieved[13–16,44]. For isotropic elastic moduli, the bending, rolling or curling behaviour can be calculated using the classical Timoshenko formula[62]. In this case, a strained bilayer NM strip tends to roll (along its long direction) into a tube when the strip is wide or into a ring when the strip is narrow. A long narrow strip may, however, form a coil, owing to shear terms in the minimization of energy. Figure 3a illustrates the conditions. A tube will form if the width, W, of the strip is large relative to the radius of curvature determined by the bilayer strain, which is related to the circumference, L_0, of the tube. Beyond a critical angle, θ_c, determined by $W/L_0 = \sin(\theta_c)$, a long, strained NM strip will roll into a coil of radius R_0, which also is related to the radius of curvature determined by the bilayer strain. More varied shapes can form when the materials have elastically soft and hard directions. Figure 3b shows rolled-up tubes of GaAs NMs with embedded quantum wells, which have applications as an unusual type of optical resonator[63]. Fig. 3c shows

spirals that form ribbon-shaped SiGe NMs[64]. The centre regions remain flat because they are attached to the underlying substrate. Lithographic patterning allows elaborate architectures that create many new design opportunities in semiconductor devices, as described in the section on applications below.

Organic nanomembranes

The chemical diversity and unique properties of organic NMs make them an attractive complement to the inorganic materials described above. Graphene provides the most compelling existing example because of its superlative mechanical, thermal and charge transport properties, and the device opportunities, unavailable to bulk graphite[65], that it affords. These observations motivate work to explore other 2D carbon allotropes, such as those with sp- and sp^2-hybridized carbon (for example graphyne and graphdiyne) with similar, molecular-scale thicknesses[7–9,66,67]. These materials, as well as other more structurally and compositionally diverse organic NMs, are expected theoretically to offer bandgaps, mobilities and other electronic characteristics that are much different from those of graphene[7–9,66,67]. For example, calculations suggest that graphdiyne has a mobility as great as $\sim 10^5$ cm^2 V^{-1} s^{-1} and a bandgap of ~ 0.5 eV (ref. 68), thereby making this type of NM attractive as a semiconductor for power-efficient, high-speed transistors, where unpatterned graphene NMs cannot be used because of their intrinsically semi-metallic nature.

The construction of 2D, ordered networks of carbon–carbon bonds in materials other than graphene remains a frontier challenge in chemistry[8,9]. Difficulties lie in the establishment of bonds with the necessary precision in high-molecular-mass systems that have the requisite solubility and avoid the propensity to aggregate[9,69]. As a result, present techniques of bulk, solution-phase synthetic chemistry typically limit the formation of organic NMs to lateral dimensions barely exceeding the molecular regime. An alternative scheme involves the assembly of molecular building blocks at interfaces, in the form of self-assembled monolayers (SAMs) or Langmuir–Blodgett films, to provide planar precursor films for NMs that form by reactions between the molecular constituents. Figure 4a illustrates the former process, and Fig. 4b shows an example of a 2,5-substituted dialkynylbenzene-bearing SAM on silica, catalytically crosslinked by alkyne metathesis to yield a highly conjugated carbon NM (a remarkably tough monolayer-thick sheet), capable of release and transfer to a silicon wafer for possible integration with established electronically active components[70]. Alternatively, crosslinking can be accomplished in related SAMs by low-energy electron bombardment[71]. For nitroaryl thiols on gold, this process forms a dense, crosslinked matrix in the aryl segments and elicits a reduction of the chain-end nitro substituents to amino groups. Detachment from the gold yields a NM that bears thiol groups on one face and amino groups on the other, with potential relevance for devices that demand different electronic interfaces on top and bottom. In Fig. 4c, the NM sits on a supporting grid; the colour arises from fluorescent labelling of the amino side. This synthesis also allows patterning by spatially modulating the electron dose, which has clear relevance for use in devices[72]. Pyrolysis yields metallic NMs that have conductivities similar to those found in graphitic forms of carbon[73], with immediate applications as conductive grids for transmission electron microscopy.

Although reported Langmuir–Blodgett and SAM techniques can yield macroscopic NMs in various forms, the absence of long-range coherence in the bonding configurations severely degrades charge transport characteristics. This challenge can be addressed with interfacial interactions templated by monocrystalline substrates. Recent work demonstrates the feasibility of this idea in the case of graphene-like NMs, in a manner that seems to be extendable to other organic NMs[74,75]. Figure 4d, e shows chevron-shaped graphene ribbons synthesized on a Au(111) crystal through the thermolytic condensation and cyclodehydrogenetic conversion of the molecular precursor 6,11-dibromo-1,2,3,4-tetraphenyltriphenylene[74]. These ribbons align and assemble along the direction of the corrugation of the herringbone reconstruction of the Au(111) substrate. Related approaches have yielded interesting classes of crystalline
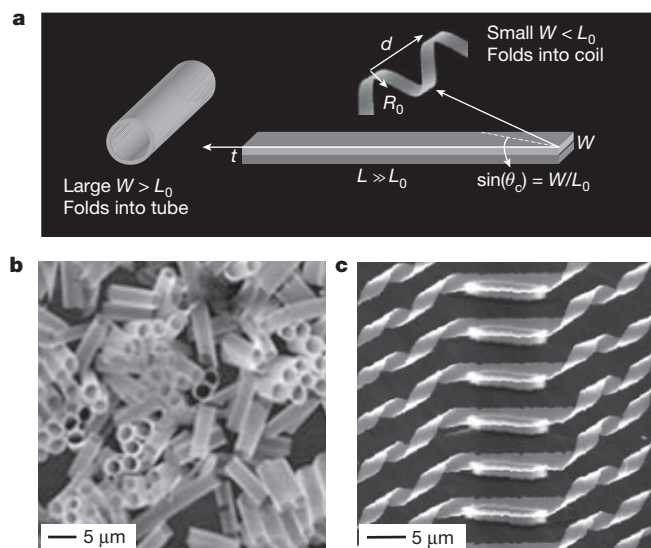
Figure 3 | Inorganic monocrystalline semiconductor NMs in non-planar configurations. a, Rolling and curling in a strained bilayer NM, illustrating the geometric parameters that determine the morphology: $L_0 = 2\pi R_0$ is the circumference of a tube that may form; L and W are the length and width of the strip, respectively; and t is its thickness. The critical angle for coil formation over tube formation is θ_c. The arrows indicate the folding direction[14]. **b**, SEM image of a collection of GaAs NMs with embedded quantum well structures. The tubular shapes form on release from the substrate, owing to strain in the epitaxial layers[63]. **c**, SEM image of an array of partly released spiral structures formed by SiGe (10 nm)/Si (7 nm)/Cr (20 nm) NMs attached at their centres to a silicon wafer[64].

monolayers from assemblies of molecules of 1,4-benzenediboronic acid, interlinked by the B–O–B bonds of the boroxine moiety[75]. Recent results suggest that other covalent organic frameworks can be synthesized using graphene as a templating substrate[76]. With additional work, these types of chemistry have great potential to yield the classes of monocrystalline organic NMs considered in this Review.
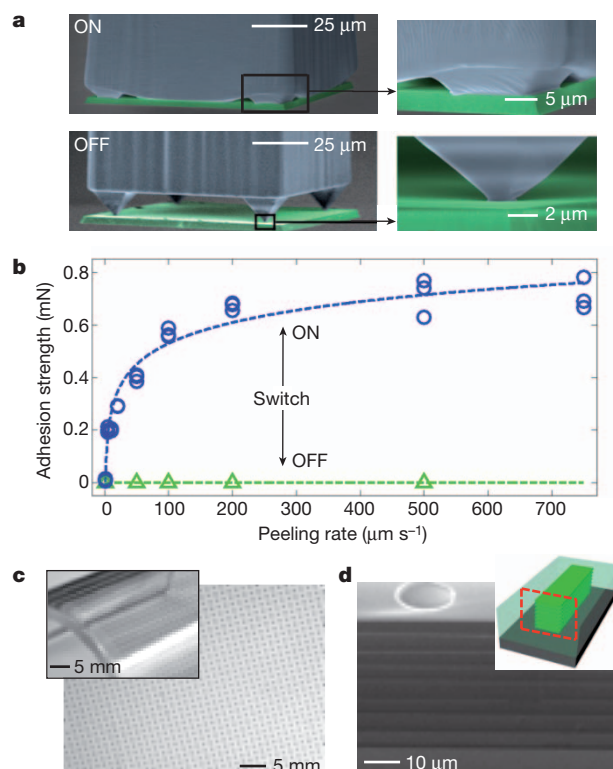
Single– and multilayer assembly

Systematic scientific studies benefit from, and engineering applications require, reliable techniques for integrating NMs into device or test structures. Materials created by bulk synthesis or by batch exfoliation from solids or surfaces can be assembled most readily through processes that start with suspensions of NMs in fluids[77–79], where adapted forms of ultracentrifugation, membrane filtration and chromatography offer means of separation, size selection and purification[80–82]. Langmuir–Blodgett techniques and controlled precipitation can yield thin-film-type assemblies, in single- or multilayer formats[77,80,81,83]. The resulting

deposits can be transferred to substrates of interest for integration into devices as thin films. The levels of control in such solution-based processes can be enhanced through patterned surface functionalization, controlled fluid flows, capillarity or shape complementarity to guide the placement of individual NMs, using schemes originally developed for nanowires/nanotubes and for small-scale integrated circuits[84,85]. The intrinsically stochastic nature of these processes of guided self-assembly, however, imposes limits on the yields and the placement accuracy.

By contrast, NMs formed at interfaces that make controlled release possible can be manipulated using purely deterministic assembly techniques, with high yields and extreme accuracy in position, orientation and layout. Such capabilities are essential for all foreseeable applications in electronics, because of the requirement to integrate the NMs at specific locations in larger systems with submicrometre accuracy. The most well-developed approaches exploit forms of printing using soft, elastomeric stamps that allow manipulation of NMs without exceeding the critical strain levels for structural damage[10–12,36,43,86]. As an advanced example, Fig. 5a shows a structured stamp (blue) designed to allow pressure-modulated adhesion with two states, strong adhesion (ON) and weak adhesion (OFF), to facilitate retrieval and printing, respectively; here, the stamp is 'inked' with a thick platelet of silicon (green). (Temporary 'carrier' films of photoresist allow manipulation of NMs
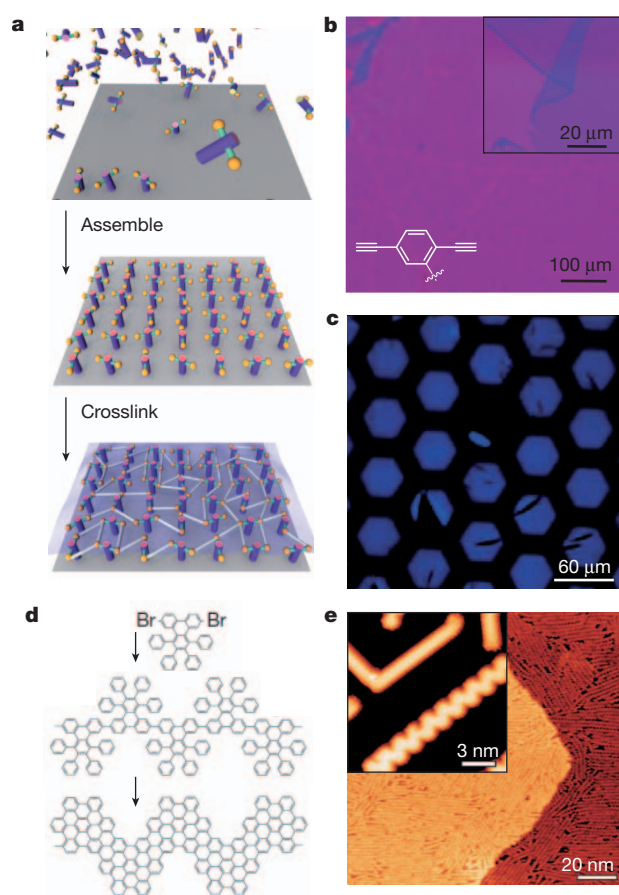
Figure 4 | **NMs of conjugated carbon and their synthesis using interfacial methods. a**, Approach to synthesis based on chemical crosslinking of a SAM[70]. **b**, Optical micrographs of a highly conjugated carbon NM synthesized by crosslinking a 2,5-substituted dialkynylbenzene SAM by alkyne metathesis, resting on a SiO_2/Si substrate[70]. A wrinkled region of the NM appears in magnified view on the top right. The chemical structure of the monomer appears on the bottom left. **c**, Fluorescence resonance energy transfer image of a ~1-nm-thick 'Janus' NM (blue) suspended over a supporting, hexagonal grid structure[71] (black). This NM, which has some tears and other defects, was formed by exposing a 4′-nitro-1,1′-biphenyl-4-thiol SAM to electrons at 100 eV and 50 mC cm^{-2}. **d**, Chemical synthesis of chevron-shaped graphene nanoribbon structures on Au(111), formed by thermolytic condensation and cyclodehydrogenetic conversion of the molecular precursor 6,11-dibromo-1,2,3,4-tetraphenyltriphenylene[74]. **e**, High-magnification (inset) and low-magnification (main image) scanning tunnelling microscope images of straight and chevron-shaped graphene nanoribbons on Au(111) synthesized in the manner illustrated in **d**[74].

Figure 5 | **Operation of elastomeric stamps for deterministic assembly of NMs, with examples of printed sparse arrays and multilayer assemblies. a**, Colourized SEM images of a single post in an elastomeric stamp (blue) that uses soft, pyramidal relief features to provide strong adhesion in a collapsed state (ON) and weak adhesion in a retracted state[10] (OFF). Control of the applied pressure allows reversible switching between these two states. **b**, Measured adhesion strength in the ON and OFF states, as a function of peeling rate[10]. Viscoelastic effects in the elastomer lead to monotonic increases in adhesion with rate, with pronounced effects observable in the ON state. The dashed lines are guides for the eye. **c**, Sparse array of GaAs membranes (small black squares) assembled by printing onto a plate of glass (main image) and a bent sheet of plastic (inset)[40]. **d**, Cross-sectional SEM of an eight-layer stack of silicon NMs (each ~340 nm thick) separated by transparent layers of polymer. Inset, schematic; the red box outlines the cross-section shown in the main image.

with arbitrarily small thicknesses.) The data in Fig. 5b illustrate this switching capability, where both geometric[10] and viscoelastic[11] effects have important roles. Printing in single or multiple cycles can populate small or large areas of flat or curved substrates with NMs at any level of coverage, from full monolayers to sparse distributions in regular or irregular layouts. Figure 5c shows an organized collection of membranes of GaAs printed onto a flat plate of glass and a bent plastic substrate, starting from a dense array released from a GaAs wafer, using techniques similar to those of Fig. 2c[40]. Recent work indicates that it is even possible to print rolled-up NMs (Fig. 3b) using soft stamps[15].

These methods, and related ones that rely on hard stamps or releasable tapes, can be used with both inorganic and organic NMs to yield various configurations, including suspended 'bridge' and 'drumhead' geometries[10,70] for resonators in nanoelectromechanical systems, large-area, continuous sheets[87] for flat-panel displays, and unusual, hybrid-material constructs[88]. As an example of the last possibility, Fig. 5d shows a stack of eight silicon NMs, each separated by a transparent polymer film, for an application in multilayer optoelectronics where these layers support waveguide arrays for phase-controlled beam steering. Fully automated tools now exist for assembling these and related structures[36,40], with printing rates of up to millions of NMs per hour, or more, depending on the sizes and layouts. Yields of nearly 100% and placement accuracy better than 1 μm can be achieved with NMs having thicknesses in the near-atomic range, relevant for single-layer NMs[89,90], and areas of square centimetres or more.

Applications in electronics and optoelectronics

The synthesis and assembly techniques described above, coupled with the unique mechanics and confinement effects provided by NMs, provide diverse capabilities in electronics. The most intriguing areas of application are those that cannot be addressed with any other bulk, thin-film or nanomaterials technology. Advanced demonstration devices of this type have been reported, perhaps most significantly in the area of flexible and stretchable electronics. In these cases, NMs of silicon[18,21,53,60,91,92], GaAs (refs 36, 40, 93) or GaN (ref. 94) serve as active materials, mounted on plastic or rubber substrates and configured in mechanical layouts and 3D nanoarchitectures that allow bending, stretching, folding, twisting and other demanding modes of deformation without inducing damage or fatigue in the materials.

An emerging NM technology that exploits these features involves intimate coupling of electronics with biological tissues in ways that would otherwise be impossible[21,45,93]. Fig. 6a shows an example of this sort of bio-integrated device, in which a thin, flexible film supports an interconnected array of 288 actively multiplexed, amplified sensing electrodes[21]. The system includes more than 2,000 silicon NM transistors on a thin sheet of polyimide, in a waterproof format that non-invasively laminates onto the epicardial surface (porcine animal model shown here), like a piece of Saran wrap. The device performs temporal and spatial mapping of electrophysiology at unprecedented levels of speed and resolution, for diagnostic purposes in surgical procedures to treat arrhythmias and other forms of cardiac disease. The colour inset in Fig. 6a shows typical data. Related components can serve as advanced surgical tools or therapeutic devices, not only in cardiology but also in neurology and other areas.

Key performance attributes of the NM transistors in these systems are enhanced by layers of SiO_2 that passivate the surfaces of the silicon to facilitate charge transport. As a result, these circuits have performance comparable to analogous systems on SOI, with normalized current outputs[60] and switching speeds[92] that exceed those of systems based on organic semiconductors or films of quantum dots or nanowires. Such NMs can in fact provide a route to plastic radio-frequency electronics[95]. Fig. 6b shows gigahertz-frequency operation in silicon NM transistors on a sheet of polyethylene terephthalate[18]. This type of technology provides a radio-frequency electronics platform of interest not only for its unusual mechanics but also for its potential as a low-cost
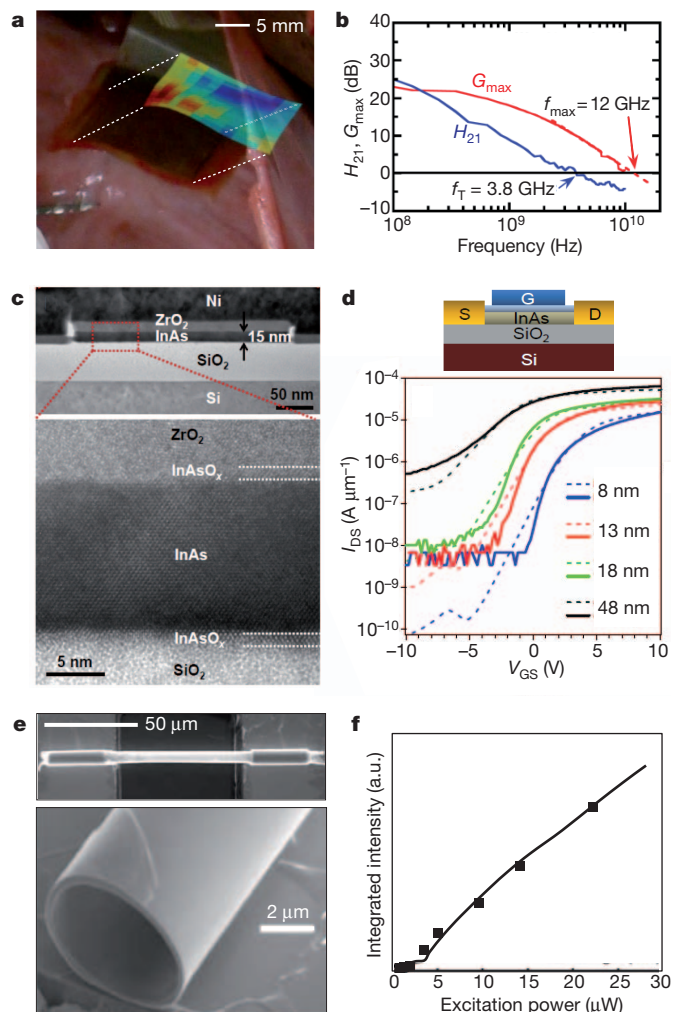
Figure 6 | NMs as active materials in unusual electronic and optoelectronic devices. a, Bio-integrated electronics for high-resolution mapping of cardiac electrophysiology in a porcine animal model, with applicability in humans[21]. The device consists of nearly three hundred independent measurement locations, with local amplifier circuits and multiplexers that collectively use more than 2,000 silicon NM transistors in a waterproof construction on a thin sheet of polyimide. The colour inset provides a representative map collected using this device. b, Small-signal current gain (H_{21}) and power gain (G_{max}) as functions of frequency for high-speed silicon NM transistors on a plastic substrate[18]. The results show limiting frequencies of 3.8 GHz for current gain (f_T) and 12 GHz for power gain (f_{max}). For power gain, the solid and dashed lines correspond to measurement and simulation, respectively. c, Cross-sectional transmission electron microscope images of a transistor that uses an InAs NM heterogeneously integrated on an oxidized silicon wafer, at moderate (top) and high (bottom) magnifications[96]. d, Measured (solid) and simulated (dashed) width-normalized drain–source currents (I_{DS}) as functions of gate–source voltage (V_{GS}) for transistors as in c with InAs NM thicknesses of 8, 13, 18 and 48 nm (ref. 96). A cross-sectional schematic of the transistor is shown at top. e, SEM images of a cylindrical microcavity laser formed with a NM (~50 nm thick, with two InGaAs/GaAs quantum dot layers and a pseudomorphic $In_{0.18}Ga_{0.82}As$ quantum well), at low (top) and high (bottom) magnifications[97]. f, Integrated output intensity as a function of excitation power (HeNe laser emission at 632.8 nm) for emission in an optically resonant mode with a wavelength of 1,240.7 nm (ref. 97).

alternative to conventional systems, which require semiconductor wafers as substrates.

In the devices of Fig. 6a, b, NMs overcome incompatibilities in processing and growth conditions between high-temperature active materials and low-temperature substrates, by exploiting the mechanics illustrated in Fig. 1a. A related advantage appears in the context of mainstream electronics in cases where integration by means of heteroepitaxial growth or wafer bonding is impossible because of mismatches in lattice constants

or coefficients of thermal expansion, respectively. For example, silicon NMs can be bonded directly to germanium substrates and vice versa, with narrow interfaces and high cross-interface electrical conductivity, which is challenging to achieve with bulk wafer-bonding techniques. Such unique capabilities make possible a range of group-IV optoelectronic and tunnelling devices in which charge transport across interfaces is essential[23].

Equally important in digital integrated circuits, but not typically requiring cross-interface transport, is joining compound-semiconductor NMs with silicon wafers, down to the level of individual devices. There the high mobilities and conductivities of InAs NMs, for instance, could help to overcome fundamental limitations on the speed and energy efficiency of silicon transistors. Figure 6c shows cross-sectional transmission electron micrographs of a transistor formed from an InAs NM printed onto a silicon wafer[96] using the techniques described in the previous section and the bonding mechanics illustrated in Fig. 1a. Figure 6d illustrates the transistor schematically and shows its current–voltage characteristics. Strong confinement effects allow orders-of-magnitude increases in switching ratios compared with bulk devices, mainly through improved electrostatic coupling to the gate, and reductions in the maximum currents due to a transition from 3D to 2D transport. Figure 6d shows trends as the thickness of the NM increases from 8 to 48 nm. Interfacial layers of $InAsO_x$, thermally grown onto the InAs NMs before device integration, are critically important in reducing the interfacial trap densities.

Passivated NMs can also be used to achieve radically different device geometries. As an example, Fig. 6e shows a cylindrical microcavity laser formed by stress-induced rolling of a GaAs/InGaAs NM with embedded, self-organized quantum dots as the gain medium[97]. As configured in a suspended state, the tube yields a cavity with excellent coupling of the maximum field intensity and the gain region. Minimal optical scattering and reduced non-radiative recombination at surface defects result from epitaxially smooth surfaces and effects of carrier confinement, respectively. The graph in Fig. 6f shows measured characteristics of the microcavity laser. Non-planar device designs represent an interesting alternative to the more widely explored microcavities based on photonic crystals, microdisks, micropillars and other geometries. These and other unique 3D nanoarchitectures allowed by NMs are also being explored for use in energy storage devices[98], sensors[59] and other components[63].

Conclusions and outlook

The existence of a recently developed, powerful set of capabilities in NM materials and assembly, taken together with multiple important and uniquely addressable application areas, provides excellent motivation for expanded activities in this rapidly emerging field. The possible topics for basic study are many, and include investigations of physical, chemical and transport properties at interfaces between heterogeneous, printed stacks of NMs; phenomena in the limit of ultrathin geometries where quantum confinement, interface depletion effects and molecular modification can be important; modified phonon and thermal characteristics for controlled heat flow in structured NMs; and strain engineering for spatially modulated bandgap properties.

For work in engineering, the most promising areas are in systems with operational features that cannot be achieved using established approaches. The overall device integration, as currently practiced, involves release of NMs from a source substrate, followed by assembly and final interconnection. This sequence is much different from the prevailing trend in conventional electronics, where individual devices, produced at the highest levels of interconnection possible on a semiconductor wafer, are diced and assembled as a terminal packaging step. The engineering challenges and balance of costs associated with the NM approach are important topics of research and development in manufacturing. An optimized process might incorporate a blend of strategies whereby, for example, some significant level of integration is accomplished on the NMs before their release and assembly, depending on the details of the application.

Techniques for synthesizing NMs are central to all future activities. New ideas are needed to expand the range of inorganic NM materials beyond those that can be achieved by known exfoliation, etching, epitaxial and bonding methods. As for other classes of nanomaterials, morphological and chemical properties of the surfaces of NMs are paramount. In some cases, existing technologies for surface passivation can be adapted to create multilayer NM structures, such as $SiO_2/Si/SiO_2$, that embed critical interfaces and isolate them from the environment. In others, these surfaces can be exposed and appropriately functionalized for applications in sensors. Progress on these and related topics will facilitate applications in fields inclusive of but wider ranging than electronics and optoelectronics, such as nanoelectromechanical devices, photonic/plasmonic structures, thermal and/or mechanical energy-harvesting elements, micro/nanoscale pressure sensors, micro/nanofluidic devices, molecular sensors, sieves, scaffolds for cell culture and others. Synthesis of functionally useful organic NMs is a persistent and notable challenge in chemistry, but one that now seems possible to address by interfacial assembly and crosslinking on monocrystalline substrates. For all classes of materials, understanding the physics of transport in shaped, chemically functionalized and/or strain-engineered NMs may lead to additional properties that lie beyond those that can be achieved otherwise.

Application opportunities seem to be particularly promising in bio-integrated systems, where many NM materials and structures might be combined in packages that establish and actively maintain intimate, dynamic interfaces with the body. In such cases, organic and inorganic NMs could function together, with the former at the bio-interface for sensing, exchanging materials and establishing bio-compatibility, and the latter separately located for the purposes of actuating, processing and transmitting data and providing power. The ability to engineer 'soft', elastic responses and 3D, curvilinear configurations in NMs with optimized nanoarchitectures on 'tissue-like' substrates will be essential to satisfying dimensional and mechanical requirements set by biological constraints. Understanding the nanomechanics of these hard–soft hybrid-material constructs, where the elastic moduli can differ by a factor of one million or more, will be necessary to allow precise physical matching to tissues. Recent advances, for example, demonstrate the ability to form NM electronics with the properties of the epidermis[99]. An ultimate goal might be systems based on man-made NMs that provide seamless, integrated functions in living systems, potentially rivalling those of naturally occurring NMs in biology. The interesting fundamental science, the diverse possibilities for creative engineering and the strong potential for broadly influential outcomes make this field of NM research a fertile one for future investigation.

1. Steigerwald, M. L. & Brus, L. E. Semiconductor crystallites — a class of large molecules. *Acc. Chem. Res.* **23,** 183–188 (1990).
2. Hebard, A. F. Buckminsterfullerene. *Annu. Rev. Mater. Sci.* **23,** 159–191 (1993).
3. Lu, W. & Lieber, C. M. Nanoelectronics from the bottom up. *Nature Mater.* **6,** 841–850 (2007).
4. Novoselov, K. S. Two-dimensional atomic crystals. *Proc. Natl Acad. Sci. USA* **102,** 10451–10453 (2005).
5. Osada, M. & Sasaki, T. Exfoliated oxide nanosheets: new solution to nanoelectronics. *J. Mater. Chem.* **19,** 2503–2511 (2009).
6. Ma, R. & Sasaki, T. Nanosheets of oxides and hydroxides: ultimate 2D charge-bearing functional crystallites. *Adv. Mater.* **22,** 5082–5104 (2010).
7. Hirsch, A. The era of carbon allotropes. *Nature Mater.* **9,** 868–871 (2010).
8. Diederich, F. & Kivala, M. All-carbon scaffolds by rational design. *Adv. Mater.* **22,** 803–812 (2010).
9. Sakamoto, J., van Heijst, J., Lukin, O. & Schlüter, A. D. Two-dimensional polymers: just a dream of synthetic chemists? *Angew. Chem. Int. Ed.* **48,** 1030–1069 (2009).
10. Kim, S. *et al.* Microstructured elastomeric surfaces with reversible adhesion and examples of their use in deterministic assembly by transfer printing. *Proc. Natl Acad. Sci. USA* **107,** 17095–17100 (2010).
11. Meitl, M. A. *et al.* Transfer printing by kinetic control of adhesion to an elastomeric stamp. *Nature Mater.* **5,** 33–38 (2005).
    Report on deterministic assembly of micro/nanoscale objects with elastomeric stamps.
12. Lee, K. J. *et al.* Large-area, selective transfer of microstructured silicon: a printing-based approach to high-performance thin-film transistors supported on flexible substrates. *Adv. Mater.* **17,** 2332–2336 (2005).
13. Cavallo, F. & Lagally, M. G. Semiconductors turn soft: inorganic nanomembranes. *Soft Matter* **6,** 439–455 (2010).

14. Huang, M. *et al.* Nanomechanical architecture of strained bilayer thin films: from design principles to experimental fabrication. *Adv. Mater.* **17,** 2860–2864 (2005).
15. Li, X. L. Strain induced semiconductor nanotubes: from formation process to device applications. *J. Phys. D Appl. Phys.* **41,** 193001 (2008).
16. Schmidt, O. G. & Eberl, K. Thin solid films roll up into nanotubes. *Nature* **410,** 168–168 (2001).
    Report on rolled-up tubes from initially flat NMs.
17. Ko, H. C. *et al.* Curvilinear electronics formed using silicon membrane circuits and elastomeric transfer elements. *Small* **5,** 2703–2709 (2009).
18. Sun, L. *et al.* 12-GHz thin-film transistors on transferrable silicon nanomembranes for high-performance flexible electronics. *Small* **6,** 2553–2557 (2010).
19. Ahn, J.-H. *et al.* Heterogeneous three-dimensional electronics by use of printed semiconductor nanomaterials. *Science* **314,** 1754–1757 (2006).
20. Yuan, H.-C. *et al.* Flexible photodetectors on plastic substrates by use of printing transferred single-crystal germanium membranes. *Appl. Phys. Lett.* **94,** 013102 (2009).
21. Viventi, J. *et al.* A conformal, bio-interfaced class of silicon electronics for mapping cardiac electrophysiology. *Sci. Transl. Med.* **2,** 24ra22 (2010).
    Report on flexible NM electronics for mapping cardiac electrophysiology.
22. Symon, K. R. *Mechanics* 3rd edn (Addison-Wesley, 1971).
23. Kiefer, A. M. *et al.* Si/Ge junctions formed by nanomembrane bonding. *ACS Nano* **5,** 1179–1189 (2011).
24. Chen, F. *et al.* Quantum confinement, surface roughness, and the conduction band structure of ultrathin silicon membranes. *ACS Nano* **4,** 2466–2474 (2010).
25. Zhang, P. *et al.* Electronic transport in nanometre-scale silicon-on-insulator membranes. *Nature* **439,** 703–706 (2006).
    Report on the influence of surfaces in controlling conduction in NMs.
26. Yu, J.-K., Mitrovic, S., Tham, D., Varghese, J. & Heath, J. R. Reduction of thermal conductivity in phononic nanomesh structures. *Nature Nanotechnol.* **5,** 718–721 (2010).
27. Tang, J. *et al.* Holey silicon as an efficient thermoelectric material. *Nano Lett.* **10,** 4279–4283 (2010).
28. Fujita, M., Takahashi, S., Tanaka, Y., Asano, T. & Noda, S. Simultaneous inhibition and redistribution of spontaneous light emission in photonic crystals. *Science* **308,** 1296–1298 (2005).
29. Siriani, D. F. *et al.* Mode control in photonic crystal vertical-cavity surface-emitting lasers and coherent arrays. *IEEE J. Sel. Top. Quantum Electron.* **15,** 909–917 (2009).
30. Coleman, J. N. *et al.* Two-dimensional nanosheets produced by liquid exfoliation of layered materials. *Science* **331,** 568–571 (2011).
31. Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V. & Kis, A. Single-layer MoS$_2$ transistors. *Nature Nanotechnol.* **6,** 147–150 (2011).
32. Davis, G. D., Viljoen, P. E. & Lagally, M. G. Determination of shallow core level spectra in selected compound semiconductors. *J. Electron Spectrosc.* **20,** 305–318 (1980).
33. Baca, A. J. *et al.* Printable single-crystal silicon micro/nanoscale ribbons, platelets and bars generated from bulk wafers. *Adv. Funct. Mater.* **17,** 3051–3062 (2007).
34. Mack, S., Meitl, M. A., Baca, A. J., Zhu, Z. T. & Rogers, J. A. Mechanically flexible thin-film transistors that use ultrathin ribbons of silicon derived from bulk wafers. *Appl. Phys. Lett.* **88,** 213101 (2006).
35. Ko, H. C., Baca, A. J. & Rogers, J. A. Bulk quantities of single-crystal silicon micro-/nanoribbons generated from bulk wafers. *Nano Lett.* **6,** 2318–2324 (2006).
36. Yoon, J. *et al.* GaAs photovoltaics and optoelectronics using releasable multilayer epitaxial assemblies. *Nature* **465,** 329–333 (2010).
37. Yablonovitch, E., Gmitter, T., Harbison, J. P. & Bhat, R. Extreme selectivity in the lift-off of epitaxial GaAs films. *Appl. Phys. Lett.* **51,** 2222–2224 (1987).
38. Konagai, M., Sugimoto, M. & Takahashi, K. High efficiency GaAs thin film solar cells by peeled film technology. *J. Cryst. Growth* **45,** 277–280 (1978).
39. Stern, F. & Woodall, J. M. Photon recycling in semiconductor lasers. *J. Appl. Phys.* **45,** 3904–3906 (1974).
40. Park, S. I. *et al.* Printed assemblies of inorganic light-emitting diodes for deformable and semitransparent displays. *Science* **325,** 977–981 (2009).
41. Sapienza, L. *et al.* Cavity quantum electrodynamics with Anderson-localized modes. *Science* **327,** 1352–1355 (2010).
42. Matthews, J. W. & Blakeslee, A. E. Defects in epitaxial multilayers: I. Misfit dislocations. *J. Cryst. Growth* **27,** 118–125 (1974).
43. Menard, E., Lee, K. J., Khang, D. Y., Nuzzo, R. G. & Rogers, J. A. A printable form of silicon for high performance thin film transistors on plastic substrates. *Appl. Phys. Lett.* **84,** 5398–5400 (2004).
44. Huang, M., Cavallo, F., Liu, F. & Lagally, M. G. Nanomechanical architecture of semiconductor nanomembranes. *Nanoscale* **3,** 96–120 (2011).
45. Rogers, J. A., Someya, T. & Huang, Y. G. Materials and mechanics for stretchable electronics. *Science* **327,** 1603–1607 (2010).
    Review of stretchable and curvilinear electronics based on shape-engineered NMs and other materials.
46. Moutanabbir, O. & Gösele, U. Heterogeneous integration of compound semiconductors. *Annu. Rev. Mater. Res.* **40,** 469–500 (2010).
47. Zhang, Y. *et al.* The fabrication of large-area, free-standing GaN by a novel nanoetching process. *Nanotechnology* **22,** 045603 (2011).
48. Kando, H. *et al.* in *Microwave Conf. Proc. (APMC), 2010 Asia-Pacific* 920–923 (IEEE, 2010).
49. Liu, Z., Wu, J., Duan, W., Lagally, M. & Liu, F. Electronic phase diagram of single-element silicon "strain" superlattices. *Phys. Rev. Lett.* **105,** 016802 (2010).
50. Euaruksakul, C. *et al.* Influence of strain on the conduction band structure of strained silicon nanomembranes. *Phys. Rev. Lett.* **101,** 147403 (2008).
51. Scott, S. A. & Lagally, M. G. Elastically strain-sharing nanomembranes: flexible and transferable strained silicon and silicon–germanium alloys. *J. Phys. D.* **40,** R75–R92 (2007).

52. Roberts, M. M. *et al.* Elastically relaxed free-standing strained-silicon nanomembranes. *Nature Mater.* **5,** 388–393 (2006).
    Report on strain engineering in NMs, to achieve unique electronic properties.
53. Yuan, H.-C., Ma, Z., Roberts, M. M., Savage, D. E. & Lagally, M. G. High-speed strained-single-crystal-silicon thin-film transistors on flexible polymers. *J. Appl. Phys.* **100,** 013708 (2006).
54. Huang, M. *et al.* Mechano-electronic superlattices in silicon nanoribbons. *ACS Nano* **3,** 721–727 (2009).
55. Paskiewicz, D. M., Tanto, B., Savage, D. E. & Lagally, M. G. Defect-free single-crystal SiGe: a new material from nanomembrane strain engineering. *ACS Nano* **5,** 5814–5822 (2011).
56. Paskiewicz, D. M., Scott, S. A., Savage, D. E., Celler, G. K. & Lagally, M. G. Symmetry in strain engineering of nanomembranes: making new strained materials. *ACS Nano* **5,** 5532–5542 (2011).
57. Sanchez-Perez, J. R. *et al.* Direct-bandgap light-emitting germanium in tensilely strained nanomembranes. *Proc. Natl Acad. Sci. USA* (submitted).
58. Huang, G. S., Mei, Y. F., Thurmer, D. J., Coric, E. & Schmidt, O. G. Rolled-up transparent microtubes as two-dimensionally confined culture scaffolds of individual yeast cells. *Lab Chip* **9,** 263–268 (2009).
59. Bernardi, A. *et al.* On-chip Si/SiO$_x$ microtube refractometer. *Appl. Phys. Lett.* **93,** 094106 (2008).
60. Kim, D.-H. *et al.* Stretchable and foldable silicon integrated circuits. *Science* **320,** 507–511 (2008).
61. Kim, D. H., Xiao, J. L., Song, J. Z., Huang, Y. G. & Rogers, J. A. Stretchable, curvilinear electronics based on inorganic materials. *Adv. Mater.* **22,** 2108–2124 (2010).
62. Timoshenko, S. Analysis of bi-metal thermostats. *J. Opt. Soc. Am.* **11,** 233–255 (1925).
63. Chun, I. S., Bassett, K., Challa, A. & Li, X. Tuning the photoluminescence characteristics with curvature for rolled-up GaAs quantum well microtubes. *Appl. Phys. Lett.* **96,** 251106 (2010).
64. Prinz, V. & Golod, S. Elastic silicon-film-based nanoshells: formation, properties, and applications. *J. Appl. Mech. Tech. Phys.* **47,** 867–878 (2006).
65. Geim, A. K. & Novoselov, K. S. The rise of graphene. *Nature Mater.* **6,** 183–191 (2007).
66. Tahara, K., Yoshimura, T., Sonoda, M., Tobe, Y. & Williams, R. V. Theoretical studies on graphyne substructures: geometry, aromaticity, and electronic properties of the multiply fused dehydrobenzo[12]annulenes. *J. Org. Chem.* **72,** 1437–1442 (2007).
67. Haley, M. M. Synthesis and properties of annulenic subunits of graphyne and graphdiyne nanoarchitectures. *Pure Appl. Chem.* **80,** 519–532 (2008).
68. Long, M., Tang, L., Wang, D., Li, Y. & Shuai, Z. Electronic structure and carrier mobility in graphdiyne sheet and nanoribbons: theoretical predictions. *ACS Nano* **5,** 2593–2600 (2011).
69. Diederich, F. Carbon scaffolding: building acetylenic all-carbon and carbon-rich compounds. *Nature* **369,** 199–207 (1994).
70. Schultz, M. J. *et al.* Synthesis of linked carbon monolayers: films, balloons, tubes, and pleated sheets. *Proc. Natl Acad. Sci. USA* **105,** 7353–7358 (2008).
71. Zheng, Z. *et al.* Janus nanomembranes: a generic platform for chemistry in two dimensions. *Angew. Chem. Int. Ed.* **49,** 8493–8497 (2010).
72. Geyer, W. *et al.* Electron-induced crosslinking of aromatic self-assembled monolayers: negative resists for nanolithography. *Appl. Phys. Lett.* **75,** 2401–2403 (1999).
73. Turchanin, A. *et al.* One nanometer thin carbon nanosheets with tunable conductivity and stiffness. *Adv. Mater.* **21,** 1233–1237 (2009).
74. Cai, J. M. *et al.* Atomically precise bottom-up fabrication of graphene nanoribbons. *Nature* **466,** 470–473 (2010).
    Report on substrate-templated synthetic routes to graphene-like nanoribbons and NMs.
75. Zwaneveld, N. A. A. *et al.* Organized formation of 2D extended covalent organic frameworks at surfaces. *J. Am. Chem. Soc.* **130,** 6678–6679 (2008).
76. Colson, J. W. *et al.* Oriented 2D covalent organic framework thin films on single-layer graphene. *Science* **332,** 228–231 (2011).
77. Stankovich, S. *et al.* Synthesis of graphene-based nanosheets via chemical reduction of exfoliated graphite oxide. *Carbon* **45,** 1558–1565 (2007).
78. Hernandez, Y. *et al.* High-yield production of graphene by liquid-phase exfoliation of graphite. *Nature Nanotechnol.* **3,** 563–568 (2008).
79. Wu, J., Pisula, W. & Müllen, K. Graphenes as potential material for electronics. *Chem. Rev.* **107,** 718–747 (2007).
80. Dikin, D. A. *et al.* Preparation and characterization of graphene oxide paper. *Nature* **448,** 457–460 (2007).
81. Xu, Y., Bai, H., Lu, G., Li, C. & Shi, G. Flexible graphene films via the filtration of water-soluble noncovalent functionalized graphene sheets. *J. Am. Chem. Soc.* **130,** 5856–5857 (2008).
82. Green, A. A. & Hersam, M. C. Solution phase production of graphene with controlled thickness via density differentiation. *Nano Lett.* **9,** 4031–4036 (2009).
83. Sasaki, D. Y., Carpick, R. W. & Burns, A. R. High molecular orientation in mono- and trilayer polydiacetylene films imaged by atomic force microscopy. *J. Colloid Interf. Sci.* **229,** 490–496 (2000).
84. Whang, D., Jin, S., Wu, Y. & Lieber, C. M. Large-scale hierarchical organization of nanowire arrays for integrated nanosystems. *Nano Lett.* **3,** 1255–1259 (2003).
85. Knuesel, R. J. & Jacobs, H. O. Self-assembly of microscopic chiplets at a liquid-liquid-solid interface forming a flexible segmented monocrystalline solar cell. *Proc. Natl Acad. Sci. USA* **107,** 993–998 (2010).
86. Menard, E., Nuzzo, R. G. & Rogers, J. A. Bendable single crystal silicon thin film transistors formed by printing on plastic substrates. *Appl. Phys. Lett.* **86,** 093507 (2005).

87. Bae, S. *et al.* Roll-to-roll production of 30-inch graphene films for transparent electrodes. *Nature Nanotechnol.* **5,** 574–578 (2010).

88. Scott, S. A., Paskiewicz, D. M., Savage, D. E. & Lagally, M. G. Silicon nanomembranes incorporating mixed crystal orientations. *ECS Trans.* **16,** 215–218 (2008).

89. Unarunotai, S. *et al.* Layer-by-layer transfer of multiple, large area sheets of graphene grown in multilayer stacks on a single SiC wafer. *ACS Nano* **4,** 5591–5598 (2010).

90. Unarunotai, S. *et al.* Transfer of graphene layers grown on SiC wafers to other substrates and their integration into field effect transistors. *Appl. Phys. Lett.* **95,** 202101 (2009).

91. Ko, H. C. *et al.* A hemispherical electronic eye camera based on compressible silicon optoelectronics. *Nature* **454,** 748–753 (2008).

92. Yuan, H.-C., Celler, G. K. & Ma, Z. 7.8-GHz flexible thin-film transistors on a low-temperature plastic substrate. *J. Appl. Phys.* **102,** 034501 (2007).

93. Kim, R.-H. *et al.* Waterproof AlInGaP optoelectronics on stretchable substrates with applications in biomedicine and robotics. *Nature Mater.* **9,** 929–937 (2010).

94. Lee, K. J. *et al.* Bendable GaN high electron mobility transistors on plastic substrates. *J. Appl. Phys.* **100,** 124507 (2006).

95. Yuan, H.-C., Qin, G., Celler, G. K. & Ma, Z. Bendable high-frequency microwave switches formed with single-crystal silicon nanomembranes on plastic substrates. *Appl. Phys. Lett.* **95,** 043109 (2009).

96. Ko, H. *et al.* Ultrathin compound semiconductor on insulator layers for high-performance nanoscale transistors. *Nature* **468,** 286–289 (2010).
   Report on transistors using NMs of InAs printed onto silicon substrates.

97. Li, F. & Mi, Z. T. Optically pumped rolled-up InGaAs/GaAs quantum dot microtube lasers. *Opt. Express* **17,** 19933–19939 (2009).

98. Ji, H.-X. *et al.* Self-wound composite nanomembranes as electrode materials for lithium ion batteries. *Adv. Mater.* **22,** 4591–4595 (2010).

99. Kim, D. H. *et al.* Epidermal electronics. *Science* **333,** 838 (2011).
   Report on NM electronics with mechanical properties matched to the epidermis, for skin-mounted devices.

Author Contributions J.A.R. led the preparation of the manuscript, wrote the sections 'Single- and multilayer assembly', 'Applications in electronics and optoelectronics' and 'Conclusions and outlook', and assembled the figures. M.G.L. and J.A.R. wrote the section 'Inorganic nanomembranes'. R.G.N. and J.A.R. wrote the section 'Organic nanomembranes'. All three authors contributed to the introduction and to editorial modifications of the overall text.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to J.A.R. (jrogers@illinois.edu).

ARTICLE

Human metabolic individuality in biomedical and pharmaceutical research

Karsten Suhre[1,2,3], So-Youn Shin[4]*, Ann-Kristin Petersen[5]*, Robert P. Mohney[6], David Meredith[7], Brigitte Wägele[1,8], Elisabeth Altmaier[1], CARDIoGRAM†, Panos Deloukas[4], Jeanette Erdmann[9], Elin Grundberg[4,10], Christopher J. Hammond[10], Martin Hrabé de Angelis[11,12], Gabi Kastenmüller[1], Anna Köttgen[13], Florian Kronenberg[14], Massimo Mangino[10], Christa Meisinger[15], Thomas Meitinger[16,17], Hans-Werner Mewes[1,8], Michael V. Milburn[6], Cornelia Prehn[11], Johannes Raffler[1,2], Janina S. Ried[5], Werner Römisch-Margl[1], Nilesh J. Samani[18], Kerrin S. Small[10], H.-Erich Wichmann[19,20,21], Guangju Zhai[10], Thomas Illig[22], Tim D. Spector[10], Jerzy Adamski[11,12], Nicole Soranzo[4]* & Christian Gieger[5]*

Genome-wide association studies (GWAS) have identified many risk loci for complex diseases, but effect sizes are typically small and information on the underlying biological processes is often lacking. Associations with metabolic traits as functional intermediates can overcome these problems and potentially inform individualized therapy. Here we report a comprehensive analysis of genotype-dependent metabolic phenotypes using a GWAS with non-targeted metabolomics. We identified 37 genetic loci associated with blood metabolite concentrations, of which 25 show effect sizes that are unusually high for GWAS and account for 10–60% differences in metabolite levels per allele copy. Our associations provide new functional insights for many disease-related associations that have been reported in previous studies, including those for cardiovascular and kidney disorders, type 2 diabetes, cancer, gout, venous thromboembolism and Crohn's disease. The study advances our knowledge of the genetic basis of metabolic individuality in humans and generates many new hypotheses for biomedical and pharmaceutical research.

Understanding the role of genetic predispositions and their interaction with environmental factors in complex chronic diseases is key to the development of safe and efficient therapies, to diagnosis and to prevention. GWAS have identified hundreds of disease-risk loci[1]; however, functional information on the underlying biological processes is often lacking[2]. Previously, we have shown the promise of using associations with blood metabolites as functional intermediate phenotypes (so-called genetically determined metabotypes (GDMs)) to understand the potential relevance of genetic variants for biomedical and pharmaceutical research[3,4]. Building on this previous work, we present here the most comprehensive evaluation of genetic variance in human metabolism so far, combining genetics and metabolomics for hypothesis generation in a GWAS. We used an extensive, non-targeted and metabolome-wide panel of small molecules, analysing >250 metabolites from 60 biochemical pathways in serum samples from 2,820 individuals from two large population-based European cohorts. We identified 37 genetic loci that were significant at a stringent genome-wide threshold. In contrast to most GWAS, these loci showed exceptionally large effect

sizes of 10–60% per allele copy in 25 loci. In the majority of cases, a protein that is biochemically related to the associated metabolic traits is encoded at these loci. As a proof-of-principle validation of new discoveries, we experimentally validated the predicted function of *SLC16A9* as a carnitine efflux transporter. We further cross-referenced these loci with databases of disease-related and pharmaceutically-relevant genetic associations, uncovering hitherto unknown links and providing new hypotheses for the functions of these loci. We have made a knowledge-base resource publically available via a web-server to aid future functional studies, and biological as well as clinical interpretation of GWAS findings. This study provides compelling evidence for novel associations of metabolic traits with a wide range of loci of biomedical and pharmaceutical interest, and indicates a powerful new paradigm for dissecting human metabolic and disease pathways.

Study design

Metabolic profiling was done on fasting serum from participants in the German KORA F4 study ($n = 1,768$) and the British TwinsUK study

[1]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany. [2]Faculty of Biology, Ludwig-Maximilians-Universität, Großhaderner Str. 2, 82152 Planegg-Martinsried, Germany. [3]Department of Physiology and Biophysics, Weill Cornell Medical College in Qatar, Education City, Qatar Foundation, PO Box 24144, Doha, State of Qatar. [4]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1HH, UK. [5]Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany. [6]Metabolon Inc., Durham, PO Box 110407, Research Triangle Park, North Carolina 27709, USA. [7]School of Life Sciences, Oxford Brookes University, Gipsy Lane, Headington, Oxford OX3 0BP, UK. [8]Department of Genome-oriented Bioinformatics, Life and Food Science Center Weihenstephan, Technische Universität München, Alte Akademie 1, 85354 Freising, Germany. [9]Universität zu Lübeck, Medizinische Klinik II, Ratzeburger Allee 160, 23538 Lübeck, Germany. [10]Department of Twin Research & Genetic Epidemiology, King's College London, St Thomas' Hospital Campus, 1st Floor South Wing Block, 4 Westminster Bridge Road, London SE1 7EH, UK. [11]Institute of Experimental Genetics, Genome Analysis Center, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany. [12]Institute of Experimental Genetics, Life and Food Science Center Weihenstephan, Technische Universität München, Alte Akademie 1, 85354 Freising, Germany. [13]Renal Division, University Hospital Freiburg, Breisacherstrasse 66, 79106 Freiburg, Germany. [14]Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Christoph Probst Platz 1, 6020 Innsbruck, Austria. [15]Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany. [16]Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany. [17]Institute of Human Genetics, Klinikum rechts der Isar, Technische Universität München, Ismaninger Straße 22, 81675 München, Germany. [18]Department of Cardiovascular Sciences, University of Leicester and Leicester NIHR Biomedical Research Unit in Cardiovascular Disease, Glenfield Hospital, University Road, Leicester LE3 7RH, UK. [19]Institute of Epidemiology I, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany. [20]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Geschwister-Scholl-Platz 1, 80539 München, Germany. [21]Klinikum Grosshadern, Marchioninistraße 15, 81377 München, Germany. [22]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany.
*These authors contributed equally to this work.
†A list of authors and their affiliations appears in Supplementary Information.

($n = 1,052$), using ultrahigh-performance liquid-phase chromatography and gas-chromatography separation, coupled with tandem mass spectrometry[5–7]. We achieved highly efficient profiling (24 min per sample) with low median process variability ($<12\%$) of more than 250 metabolites, covering more than 60 biochemical pathways of human metabolism (Supplementary Table 1). On the basis of our previous observation that ratios between metabolite concentrations can strengthen the association signal and provide new information about possible metabolic pathways[4,8], we included all pairs of ratios between these metabolites in the genome-wide statistical analysis. To reduce the computational and data-storage burden associated with meta-analysing more than 37,000 metabolites and ratios, we applied a staged approach for selection of promising association signals (Supplementary Fig. 1). In the initial screening stage, we assessed the association of approximately 600,000 genotyped single nucleotide polymorphisms (SNPs) with more than 37,000 metabolic traits (concentrations and their ratios) by fitting linear models separately in both cohorts to log-transformed metabolic traits and adjusting for age, gender and family structure (Supplementary Fig. 2 and Supplementary Table 2). Next, we selected all association signals showing suggestive evidence of association with a metabolic trait in both cohorts ($P < 10^{-6}$ in both cohorts, or $P < 10^{-3}$ in one and $P < 10^{-9}$ in the other). For each of these loci, we then reassessed the association signals through fixed-effects inverse variance meta-analysis of the two cohorts for all 37,000 available traits, using imputed SNPs relative to HapMap2 data (see Methods for details). The combination of SNP and trait that yielded the smallest P value in this meta-analysis was finally selected for each locus. To account for multiple testing, we applied conservative Bonferroni correction, leading to an adjusted threshold for genome-wide significance of $P < 2.0 \times 10^{-12}$.

Study results

We identified 37 independent loci that reached genome-wide significance in the meta-analysis (Table 1 and Supplementary Tables 3 and 4). Twenty-three of these loci describe new genetic associations with metabolic traits, and 14 replicate and extend our knowledge of known GDMs, including 10 from our own studies[3,4]. We used information on the locations of SNPs in genes, on known gene functions and on regional association plots (Supplementary Fig. 2) to prioritize plausible candidate genes within associated loci. In most cases, our annotation was further supported by a statistical analysis of the association of gene relationships in published literature[9] (Supplementary Table 5). Associations with additional metabolic traits at the 37 loci presented in Table 1 may capture further biochemical information, and are provided as Supplementary Table 6. At 30 loci, the sentinel SNP mapped to a protein that was biochemically linked to the associating metabolites, for instance because it was responsible for their synthesis, degradation or metabolism. Next, we searched literature and databases extensively (see web links in Methods) to identify which of these 37 loci were previously reported as being associated with a clinical endpoint, a medically relevant intermediate phenotype, or a pharmacogenetic effect. Associations of metabolites with disease loci can be used to gain novel information about possible metabolic changes associated with biological processes underlying that association (Fig. 1, Table 1 and Supplementary Table 7). In 15 cases, such a relationship could be identified on the basis of an association of the lead SNP or a proxy ($r^2 \geq 0.8$) with the disease-associated SNPs, including those for cardiovascular disease, kidney disease, Crohn's disease, gout, cancer, adverse reactions to drug therapy and predisposing risk factors for diabetes and cardiovascular disease. In all except three loci, the SNPs are common, with minor allele frequencies greater than 10%. In 25 cases, the effect size per allele copy is larger than 10%, and up to 60% in the case of the acyl-CoA dehydrogenase (*ACADS*) locus.

Overlap with chronic disease loci

Many genetic-risk loci for heart disease, kidney failure, diabetes and other complex disorders have been identified by GWAS. However, the aetiology of these common diseases is complex and testable hypotheses are needed to develop new avenues for diagnosis and therapy. Associations of known disease-risk loci with metabolic traits allow the identification of new and potentially relevant biological processes and pathways. Here we report some examples from our study that illustrate this idea. The full association data set is freely available for further analysis and reference at http://www.gwas.eu.

Detoxification and kidney failure

N-acetylation is an important mechanism to detoxify nephrotoxic medications and environmental toxins. A reduced ability to detoxify such substances could lead to impaired kidney function. A key GDM is the N-acetyltransferase 8 (*NAT8*) locus, which was reported to associate with kidney function[10,11]. Here we found a highly significant association of variation at the *NAT8* locus with N-acetylornithine. Using this information, we investigated whether N-acetylornithine concentrations were associated with kidney function. In both our studies, we found a clear association with estimated glomerular filtration rate (eGFR), whereby higher levels of N-acetylornithine were correlated with lower eGFR ($P_{\text{KORA}} = 7.6 \times 10^{-4}$, $P_{\text{TwinsUK}} = 3.6 \times 10^{-8}$ after adjusting for age and gender). In accord with the genetic effect of the *NAT8* polymorphism in chronic kidney disease (CKD), the risk allele identified here was associated with higher N-acetylornithine concentrations. Although causality cannot be inferred from this kind of association study, the role of ornithine acetylation in the aetiology of CKD warrants further exploration.

Diabetes

Glucokinase (hexokinase 4) regulator (*GCKR*) is a major pleiotropic risk locus associated with diabetes- and cardiometabolic-related traits, such as fasting glucose and insulin levels[12], triglyceride levels[13] and CKD[11]. Here we identified a highly significant association of this locus with mannose:glucose ratios. The fasting level of mannose is lower in carriers of the risk allele, as opposed to glucose being higher. Notably, we also observed a 3.3% increase in lactate concentration per copy of the risk allele at the same locus. Little is known about the physiological role of mannose, other than its use in protein glycosylation. Mannose enters the cell via a specific transporter that is insensitive to glucose[14], and hepatic glycogen breakdown is implicated in the maintenance of plasma mannose concentrations[15]. These observations and the association with *GCKR* observed here, which is even stronger than that of glucose with *GCKR*, indicate a need for further investigation of the role of mannose as a differential biomarker, or even as a point of intervention in diabetes care.

Venous thromboembolism

With the mass-spectrometry method used here, different forms of the abundant fibrinogen A-α peptides can be detected. Fibrinogen has a role in the formation of blood clots. Its active form, the fibrinogen A-α chain ADSGEGDFXAEGGGVR, can be phosphorylated at serine 3 to ADpSGEGDFXAEGGGVR[16]. The ratio between the concentrations of these fibrinogen A-α peptides provides a measure for fibrinogen A-α phosphorylation (FAαP). Increased levels of FAαP have been observed under various physiological and pathophysiological conditions[17]. Here, three loci (*ABO*, *ALPL* and *FUT2*) associated with FAαP. Notably, these three genes are functionally linked: *ABO* (encoding ABO blood group (transferase A, α-1-3-N-acetylgalactosaminyltransferase; transferase B, α-1-3-galactosyltransferase)) and *FUT2* (encoding fucosyltransferase 2) are involved in determining the blood group, and the *ABO* locus is associated with blood levels of the alkaline phosphatase ALPL[18]. The association of *ALPL* with FAαP may be explained either by a genotype-dependent dephosphorylation of fibrinogen by ALPL, or by a genotype-dependent change in the phosphorus pool available for FAαP. Variants in the *ABO* gene are associated with many different outcomes, including venous thromboembolism (VTE)[19]. The association of *ABO* with FAαP, and thus with modified blood coagulation properties, provides a

Table 1 | Thirty-seven loci that displayed genome-wide significance in the meta-analysis

Locus & SNP id	Metabolic trait	P value	Relationship between gene function and the associated metabolic traits	Biomedical and pharmaceutical interest
ACADS rs2066938	Butyrylcarnitine/propionylcarnitine	$< 4.4 \times 10^{-305}$	Butyrylcarnitine$^+$ and propionylcarnitine$^+$ are substrates/products of ACADS	ACADS is a key enzyme in mitochondrial fatty acid β-oxidation
NAT8 rs13391552	N-acetylornithine	5.4×10^{-252}	N-acetyltransferase function of NAT8 matches the associating metabolite N-acetylornithine$^+$	Association with **glomerular filtration** and **CKD**; association of N-acetylornithine$^+$ with eGFR in this study
FADS1 rs174547	1-arachidonoylglycerophosphoethanolamine/ 1-linoleoylglycerophosphoethanolamine	8.5×10^{-116}	FADS1 substrate/product pair ratio arachidonate (20:4n6)$^+$/dihomo-linolenate (20:3n3 or n6)$^+$ is among the top associations	Association with **LDL cholesterol, HDL cholesterol and triglycerides, fasting glucose and homeostatic model assessment B (HOMA-B) Crohn's disease** and **resting heart rate**
UGT1A rs887829	Bilirubin (E,E)/oleoylcarnitine	2.9×10^{-74}	Bilirubin$^+$ is a substrate of UGT1A1	Association with **hyperbilirubinaemia**; low serum concentrations of bilirubin associate with increased risk of CAD; a SNP in *UGT1A1* is a pharmacogenetic risk factor for irinotecan toxicity
ACADM rs211718	Hexanoylcarnitine/oleate (18:1n9)	2.2×10^{-71}	Hexanoylcarnitine$^+$ is a substrate of ACADM	ACADM is a key enzyme in mitochondrial fatty acid β-oxidation
OPLAH rs6558295	5-oxoproline	1.5×10^{-59}	5-oxoproline$^+$ is a substrate of 5-oxoprolinase OPLAH	
SCD rs603424	Myristate (14:0)/myristoleate (14:1n5)	2.9×10^{-57}	SCD catalyses the Δ-9-desaturation of fatty acids, such as myristate (14:0)$^+$ to myristoleate (14:1n5)$^+$ and palmitate (16:0)$^+$ to palmitoleate (16:1n7)$^+$	Palmitoleate (16:1n7) is a lipokine linking adipose tissue to systemic metabolism
GCKR rs780094	Glucose/mannose	5.5×10^{-53}	GCKR has a role in glucose homeostasis; strong association with mannose$^+$ to glucose$^+$ ratios matches the gene's function	Association with **type 2 diabetes, fasting glucose, fasting insulin; serum uric acid; triglyceride levels; C-reactive protein; serum creatinine (eGFRcrea), Crohn's disease** and **hypertriglyceridaemia**
NAT2 rs1495743	1-methylxanthine/4-acetamidobutanoate	1.7×10^{-40}	4-acetamidobutanoate$^+$, 1-methylxanthine$^+$ and 1-methylurate$^+$ are linked to NAT2 in the xenobiotics pathways	Association with **triglyceride levels and CAD; bladder cancer** and toxicities to docetaxel and thalidomide treatment
CYP3A4 rs17277546	Androsterone sulphate	8.7×10^{-40}	CYP3A cytochrome P450 proteins metabolise androsterone sulphate$^+$	Genetic variation in androsterone metabolism is linked to the incidence of prostate cancer
ABO rs612169	ADpSGEGDFXAEGGGVR/ADSGEGDFXAEGGGVR	9.1×10^{-40}	Polymorphisms in ABO determine the blood group; association with fibrinogen peptide phosphorylation$^+$; additive effect on fibrinogen A-α phosphorylation together with FUT2 and ALPL	Association with **blood alkaline phosphatase level; pancreatic cancer; venous thromboembolism** and **phytosterol levels**
SLC2A9 rs4481233	Urate	5.5×10^{-34}	SLC2A9 (GLUT9) transports uric acid$^+$	Association with **gout**; several SNPs in *SLC2A9* associate with etoposide IC$_{50}$
CYP4A rs9332998	10-nonadecenoate (19:1n9)/10-undecenoate (11:1n1)	5.1×10^{-32}	Cytochrome P450, family 4, subfamily A, are fatty acid ω-hydroxylases; 10-undecenoate (11:1n1)$^+$ is biochemically related to ω-hydroxylated C10 fatty acids	Possible role in the etiology of hepatic steatosis, in interaction with stearoyl-coenzyme A desaturase 1
CPS1 rs2216405	Glycine	1.6×10^{-27}	Association with glycine$^+$ and creatine$^+$; creatine is produced from glycine; glycine is metabolically related to carbamoyl phosphate, which is the product of CPS1 and the entry point of ammonia into the urea cycle	Metabolomics data indicates that this association is related to a perturbed ammonia metabolism
LACTB rs2652822	Succinylcarnitine	7.2×10^{-27}	Association with succinylcarnitine$^+$; perturbed hepatic gene expression in transgenic *LACTB* mice indicates a role of LACTB in the butanoate/succinate$^+$ pathway	*LACTB* transgenic mice are obese
SLC22A1 rs662138	Isobutyrylcarnitine	7.3×10^{-25}	SLC22A1 (OCT1) translocates a broad array of organic cations; possibly also isobutyrylcarnitine$^+$ or related metabolites	Genetic variations in the *SLC22A1/SLC22A2* region are determinants of metformin pharmacokinetics
SLCO1B1 rs4149081	Eicosenoate (20:1n9 or 11)/tetradecanedioate	2.8×10^{-22}	SLCO1B1 (OATP2, OATP-C) is an organic anion transporter	Common variants in *SLCO1B1* are strongly associated with an increased risk of **statin-induced myopathy**
FUT2 rs503279	ADpSGEGDFXAEGGGVR/ ADSGEGDFXAEGGGVR	4.3×10^{-20}	FUT2 is involved in the creation of a precursor of an H antigen, and has an additive effect on fibrinogen A-α phosphorylation together with ABO and ALPL	Association with **vitamin B12 levels, total cholesterol** and **Crohn's disease**; vitamin B12 deficiency is associated with cognitive decline, cancer and CAD
ACE rs4329	Aspartylphenylalanine	8.2×10^{-20}	Angiotensin I converting enzyme (peptidyl-dipeptidase A) 1 is associated with the dipeptide aspartylphenylalanine$^+$	Association with **angiotensin-converting enzyme activity**; potential genetic interaction with *KLKB1* locus

Table 1 | Continued

Locus & SNP id	Metabolic trait	*P* value	Relationship between gene function and the associated metabolic traits	Biomedical and pharmaceutical interest
PHGDH rs477992	Serine	2.6×10^{-14}	PHGDH catalyses the first and rate-limiting step in the phosphorylated pathway of serine$^+$ biosynthesis	
ENPEP rs2087160	ADpSGEGDFXAEGGGVR/DSGEGDFXAEGGGVR	6.5×10^{-13}	ENPEP (APA, aminopeptidase A) is an amino-terminal amino peptidase; association with ratios between fibrinogen A-α peptide ADSGEGDFXAEGGGVR$^+$ and its N-terminal cleaved form DSGEGDFXAEGGGVR$^+$ indicate that fibrinogen is a substrate of ENPEP	ENPEP has a role in the catabolic pathway of the renin-angiotensin system, and regulates blood pressure; association with **blood pressure** in Asian population
AKR1C rs2518049	Androsterone sulphate/epiandrosterone sulphate	6.7×10^{-13}	AKR1C isoforms have a role in androgen$^+$ metabolism	AKR1C has a role in the etiology of cancers including prostate, brain, breast, bladder and leukaemia; potential target of jasmonates in cancer cells
NT5E rs494562	Inosine	7.4×10^{-13}	Inosine$^+$ is a substrate of the 5′-nucleotidase NT5E	NT5E is involved in purine salvage
PRODH rs2023634	Proline	2.0×10^{-22}	PRODH catalyses the first step in proline$^+$ degradation	
HPS5 rs2403254	α-hydroxyisovalerate	1.0×10^{-20}	α-hydroxyisovalerate$^+$ is found in urine of patients with phenylketonuria; phenylalanine is required for melatonin biosynthesis	Melatonin homeostasis is deranged in patients with loss of *HPS* genes (albinism)
ALPL rs10799701	ADpSGEGDFXAEGGGVR/DSGEGDFXAEGGGVR	2.9×10^{-20}	ALPL is a phosphatase and associates with A-α fibrinogen phosphorylation$^+$; it has an additive effect on fibrinogen A-α phosphorylation together with ABO and of FUT2	
SLC7A6 rs6499165	Glutaroyl carnitine/lysine	9.8×10^{-19}	Glutaryl-CoA$^+$ is an intermediate in the metabolism of lysine$^+$ and tryptophan	Deficiencies in glutaryl-CoA dehydrogenase are linked to metabolic disorders
KLKB1 rs4253252	Bradykinin, des-arg(9)	6.6×10^{-18}	Kallikrein B, plasma (Fletcher factor) 1; kallikrein-kininogen complex binds to cell surface receptors leading to the targeted action of bradykinin$^+$	**Association of bradykinin$^+$ with hypertension** confirmed in this study; potential genetic interaction with ACE locus
GLS2 rs2657879	Glutamine	3.1×10^{-17}	GLS2 catalyses the hydrolysis of glutamine$^+$	
PDXDC1 rs7200543	1-eicosatrienoylglycerophosphocholine/ 1-linoleoylglycerophosphocholine	4.5×10^{-16}	Association with the 1-eicosadienoyl- to 1-eicosatrienoyl-glycerophosphocholine$^+$ ratio indicates a role of PDXDC1 in the metabolism of C20:2 and C20:3 fatty acids	Association with **body height**
SLC22A4 rs272889	Isovalerylcarnitine	7.4×10^{-16}	SLC22A4 (OCTN1) transports isovalerylcarnitine$^+$	Association with **body height**
AHR rs12670403	Caffeine/quinate	4.8×10^{-15}	AHR is a transcription factor for CYP1A1, which metabolises caffeine$^+$	
ETFDH rs8396	Decanoylcarnitine	5.5×10^{-15}	Decanoylcarnitine$^+$ is used for energy production via β-oxidation to the electron transfer complex	ETFDH is a key enzyme in mitochondrial fatty acid β-oxidation
ELOVL2 rs9393903	Docosahexaenoate (DHA; 22:6n3)/eicosapentaenoate (EPA; 20:5n3)	1.7×10^{-14}	EPA (20:5n3)$^+$ is a substrate of ELOVL2; DHA (22:6n3)$^+$ is related to its product by a single desaturation reaction	
SLC16A9 rs7094971	Carnitine	3.4×10^{-14}	SLC16A9 (MCT9) transports free carnitine$^+$ (shown in this paper)	
IVD rs10518693	3-(4-hydroxyphenyl)lactate/isovalerylcarnitine	1.1×10^{-13}	Isovalerylcarnitine$^+$ is a transport form of isovalerate, which is the substrate of isovaleryl coenzyme A dehydrogenase (IVD)	IVD is a key enzyme in mitochondrial fatty acid β-oxidation
SLC16A10 rs7760535	Isoleucine/tyrosine	1.4×10^{-12}	SLC16A10 encodes the T-type amino acid transporter-1 (TAT1); this transporter transports tyrosine$^+$ and phenylalanine$^+$	

The metabolic trait with the strongest association at the discovery stage in both studies is reported, together with the SNP identifier and the *P* value of association from the meta-analysis. Full association data are available in Supplementary Tables 3 & 5 and at http://www.gwas.eu. The loci are labelled by the gene that is considered most likely to carry the causative SNP. Where the metabolic trait is associated with a nearby gene's function, details are provided in the column labelled 'Relationship between gene function and the associated metabolic traits'. Overlaps with associations from other GWAS studies are highlighted in bold ($R^2 > 0.8$, details are in Supplementary Table 6). $^+$, Metabolic traits that are associated with the SNP at the corresponding locus. Further information and full bibliographic references are presented in Supplementary Table 4.

functional explanation for the reported association of *ABO* with VTE risk. Moreover, if FAαP is at the basis of VTE, then *FUT2* and *ALPL* should also be investigated as VTE risk genes. This hypothesis can now be tested in the respective patient groups.

Coronary artery disease

We have shown previously[4] that strong associations with metabolic traits, derived from GWAS, can point to interesting associations with

clinical endpoints that would not otherwise be considered relevant. A recent meta-analysis with lipid traits[20], using a similar strategy, identified several genetic loci that were found to affect the risk of coronary artery disease (CAD) in the CARDIoGRAM study[21]. Six of these loci are also reported here (*ABO*, *NAT2*, *CPS1*, *NAT8*, *ALPL* and *KLKB1*), although some of them showed only weak evidence for association with CAD ($P < 0.01$) in the CARDIoGRAM study (Supplementary Table 8). Although the links are not statistically strong, the

a

Lipids

ACADS Butyryl-carnitine/propionyl-carnitine

SCD C16:0 / C16:1 C14:0 / C14:1

IVD Isovaleryl-carnitine

ETFDH Decanoyl-carnitine

Carbohydrates

GCKR Glucose / mannose

FADS1 C20:4 / C20:3

ACADM Hexanoyl-carnitine

LACTB Succinyl-carnitine

HPS5 α-hydroxy-isovalerate

ELOVL2 C22:6n3/ C20:5n3

PRODH Proline

Amino acids

PDXDC1 Eicosatri-enoyl-GPC/ 1-linoleoyl-GPC

PHGDH Serine

ALPL Fibrinogen A-α phospho-rylation

GLS2 Glutamine

ENPEP Fibrinogen A-α cleavage

Human metabolic individuality

CPS1 Glycine

ABO Fibrinogen A-α phospho-rylation

OPLAH Oxoproline

FUT2 Fibrinogen A-α phosph-orylation

SLC16A10 Isoleucine/tyrosine

ACE Aspartyl-phenyl-alanine

NT5E Inosine

Nucleo-tides

KLKB1 Bradykinin

SLC7A6 Glutaroyl carnitine/lysine

Poly-peptides

NAT2 Methyl-xanthine / acetamido-butanoate

SLC2A9 Urate

UGT1A Bilirubin

AHR Caffeine / quinate

SLC22A1 Isobutyryl-carnitine

SLC16A9 Carnitine

NAT8 N-acetyl-ornithine

CYP3A4 Androsterone sulphate

SLC22A4 Isovaleryl-carnitine

AKR1C Androsterone sulphate/epiandro-sterone sulphate

SLCO1B1 Eicosenoate/tetradecane-dioate

Metabolite transport

Secondary metabolites and xenobiotics

CYP4A C19:1n9 / C11:1n1

b

ACADS Butyryl-carnitine/propionyl-carnitine

IVD Isovaleryl-carnitine

ETFDH Decanoyl-carnitine

SCD C16:0 / C16:1 C14:0 / C14:1

ACADM Hexanoyl-carnitine

LACTB Succinyl-carnitine

GCKR Glucose / mannose

FADS1 C20:4 / C20:3

ELOVL2 C22:6n3/ C20:5n3

PRODH Proline

HPS5 α-hydroxy-isovalerate

PDXDC1 Eicosatri-enoyl-GPC/ 1-linoleoyl-GPC

PHGDH Serine

ALPL Fibrinogen A-α phospho-rylation

Loci associated with traits of biomedical and pharmaceutical interest

GLS2 Glutamine

ENPEP Fbrinogen A-α cleavage

CPS1 Glycine

ABO Fibrinogen A-α phospho-rylation

OPLAH Oxoproline

FUT2 Fibrinogen A-α phosph-orylation

SLC16A10 Isoleucine/tyrosine

ACE Aspartyl-phenyl-alanine

NT5E Inosine

KLKB1 Bradykinin

SLC7A6 Glutaroyl carnitine/lysine

NAT2 Methyl-xanthine / acetamido-butanoate

SLC2A9 Urate

UGT1A Bilirubin

AHR Caffeine / quinate

SLC22A1 Isobutyryl-carnitine

SLC16A9 Carnitine

NAT8 N-acetyl-ornithine

CYP3A4 Androsterone sulphate

SLC22A4 Isovaleryl-carnitine

AKR1C Androsterone sulphate/epiandro-sterone sulphate

SLCO1B1 Eicosenoate/tetradecane-dioate

CYP4A C19:1n9 / C11:1n1

biochemical function of the associated metabolic traits identified here may support a possible role in heart disease. For example, *NAT8* may be linked to CKD via ornithine acetylation (see above). *KLKB1*, encoding kallikrein B plasma (Fletcher factor) 1, controls blood pressure via the bradykinin pathway. In this study, a genetic variant in *KLKB1* was associated with bradykinin concentrations and we also confirmed the expected directional association of bradykinin with hypertension in both our studies ($P_{KORA} = 1.7 \times 10^{-9}$, $P_{TwinsUK} = 0.0495$, with the covariates age and gender). *ABO* and *ALPL* associated with FAαP, and we therefore speculate that genetically determined differences in FAαP and resulting blood-coagulation properties may be the basis of these associations with CAD. Furthermore, our associations indicate

Figure 1 | Genetic basis of human metabolic individuality and its overlap with loci of biomedical and pharmaceutical interest. More than 100 years ago, it was realised that inborn errors in human metabolism were 'merely extreme examples of variations of chemical behaviour which are probably everywhere present in minor degrees' and that this 'chemical individuality (confers) predisposition to and immunities from the various mishaps which are spoken of as diseases'[36]. The 37 genetically determined metabotypes (GDMs) that we have reported here explain a highly relevant amount of the total variation in the studied population and therefore contribute substantially to the genetic part of human metabolic individuality. **a**, GDMs are shown colour-coded by general metabolic pathways, together with selected associated metabolic traits, highlighting the relationship between gene function and the associated metabolic trait (see column 4 in Table 1). **b**, GDMs colour-coded by overlap with associations in previous GWAS with disease (red), intermediate risk factors for disease (yellow) and other traits (green). Locus overlap is defined here by the lead SNP reported in the national human genome research institute (NHGRI) GWAS catalogue being highly correlated ($R^2 \geq 0.8$) with the most associated SNP in the metabolomics scan (see column 5 in Table 1 and Supplementary Table 7). Note that the overlap between the metabolomics loci and the loci reported by the NHGRI GWAS catalogue is highly significant when compared to a draw of 37 randomly selected SNPs with similar properties ($P < 3 \times 10^{-6}$, see Methods).

that the role of FAαP as a biomarker for acute myocardial infarction, and the combined additive genetic effect of the *ABO*, *ALPL* and *FUT2* loci (Supplementary Fig. 4) on CAD risk, should be investigated in greater detail.

New biological and functional insights

Genome-wide association studies merely uncover statistically significant associations, and can therefore only generate biological hypotheses. Although providing experimental validation of all associations is beyond the scope of a single study, we nevertheless attempted to show that, in principle, validation is possible. The association of SNP rs7094971 in solute carrier family 16, member 9

Figure 2 | Experimental evidence for SLC16A9 (MCT9) as a carnitine efflux transporter. **a**, **b**, When 4.6 nl of [³H]-carnitine was injected into *Xenopus* oocytes, followed by incubation in medium for 90 min, efflux was significantly higher in oocytes expressing *MCT9* than in the non-injected (NI, **a**) or water-injected (WI, **b**) controls. By contrast, when oocytes were incubated in medium containing [³H]-carnitine ($4\,\mu\text{Ci ml}^{-1}$), there was no significant uptake, indicating that MCT9 does not mediate carnitine uptake (data not shown). Because some previously characterized monocarboxylic acid transporters are proton-coupled[37], the experiments were conducted at both pH_{out} 7.4 and pH_{out} 5.5, but no significant difference was observed (**a** and data not shown). In agreement with this, external unlabelled carnitine was unable to trans-stimulate [³H]-carnitine efflux, with no significant difference in efflux between MCT9-expressing oocytes in the absence or presence of 5 mM carnitine (MCT9 versus MCT9 + carn, **b**). Data are means ± s.e.m. of 6–10 oocytes per data point from 2 oocyte preparations. The y-axes represent remaining [³H]-carnitine levels (c.p.m. per oocyte). Statistical significance was determined by the Student's t-test. These results are consistent with MCT9 acting as a unidirectional carnitine efflux system when expressed in *Xenopus* oocytes. Additional experiments are required to establish the full substrate specificity of MCT9. If future studies show an appropriate cellular distribution, MCT9 could be responsible for carnitine efflux across the basolateral membrane of absorptive epithelial cells, after absorption via the well-characterized SLC22 (also known as OCTN) family of apical epithelial proton-coupled carnitine transporters[38].

(*SLC16A9*, also known as *MCT9*) with carnitine indicated that this metabolite is the substrate of this hitherto uncharacterized monocarboxylic acid transporter. We therefore tested [³H]-carnitine uptake by *SLC16A9*-expressing *Xenopus* oocytes. As shown in Fig. 2, our data show that SLC16A9 is a pH-independent carnitine efflux transporter, possibly responsible for carnitine efflux from absorptive epithelia into the blood.

Another prominent example is the highly significant association of increased urate levels, and their clinical complication of gout, with variants in the *SLC2A9* gene[22]. The association between *SLC2A9* variants and urate levels was also observed here. Although it was previously annotated as a glucose transporter, *SLC2A9* was later shown[23] to encode a high-capacity urate transporter. Similar characterization experiments by specialists in the related fields should be motivated and guided by our association data. Among the 37 GDMs reported here, we suggest that the associations with coarsely-characterized genes for enzymes and transporters that are known disease-risk loci may warrant further experimental investigation, for instance in experiments using isotope-labelled derivatives of the associated metabolites reported here as putative target substrates. We deem *NAT8* to be a prime candidate for such a study.

Pharmacogenomics

Using the pharmacogenomics knowledge base[24], we identified six GDMs as being previously associated with toxicity or adverse reactions to medication. Noteworthy are polymorphisms in the *NAT2* and *CYP4A* loci that associated with toxicities to docetaxel and thalidomide treatment[25]; the *UGT1A* locus with irinotecan toxicity[26], *SLC2A9* with the IC_{50} of etoposide[27], *SLC22A1* with metformin pharmacokinetics[28,29] and *SLCO1B1* with statin-induced myopathy[30]. In all cases, our associations with metabolic traits at these loci provide a possible novel biochemical basis for the genotype-dependent reaction to drug treatment, such as the association of *SLCO1B1* with a series of fatty acids, including tetradecanedioate and hexadecanedioate. This information can be used to support the redesign of the respective drug molecules to avoid adverse reactions. Moreover, systematic inclusion of biochemically relevant GDMs as candidate SNPs during drug trials may permit early identification of potentially adverse pharmacogenetic effects. This applies specifically to *AKR1C*, which is a novel target of jasmonates in cancer cells[31]. We reported a GDM associated with *AKR1C* which has a large effect-size on androgen metabolism. The influence of SNP rs2518049 in *AKR1C* on the efficiency and potential side effects of jasmonates should therefore be assessed in future clinical trials.

Discussion

Owing to their large effect-size and high explained variance, the 37 GDMs reported in this study indicate key genetic loci underpinning differences in human metabolism. Inclusion of these genetic variants in the statistical analysis of pre-clinical and clinical studies may facilitate identification of genotype-dependent outcomes, such as disease complications and adverse drug reactions. In two cases, we could establish a direct functional link, supported by both our studies, between a genetic variant, an intermediate metabolic trait and a disease-relevant endpoint: *KLKB1* with bradykinin and hypertension, and *NAT8* with N-acetylornithine and eGFR. We note that by discussing only associations that are supported by two independent studies at genome-wide significance, we have chosen to take a very conservative approach. On the basis of QQ-plots and coarse assumptions, we estimate that more than 500 loci with signals of association below that conservative threshold may be confirmed as GDMs in more highly powered studies in the future. Technically, it is of note that by using a single study to profile 2,820 individuals metabolically, using only 100 µl of blood serum, we replicated in this study a wide series of findings from previous large GWAS with quantitative traits, including serum levels of fasting glucose[12], bilirubin[32,33], urate[34] and

dehydroisoandrosterone sulphate[35]. Our study shows how GWAS with intermediate traits that are close to the underlying biological processes can provide new functional insights into associations from GWAS with the endpoints of complex chronic disease and drug toxicity. Future GWAS that combine multiple 'omics' technologies in a single study, including transcriptomics, proteomics, metabolomics and recent technologies for determining epigenetic modifications on a genome-wide scale, are likely to be the next big step towards a full understanding of the interaction between genetic predispositions and environmental factors in the development of complex chronic diseases, their diagnosis, prevention and safe and efficient therapy.

METHODS SUMMARY

The full Methods section provides information about study design, genetic and metabolic data collection, and data analysis.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).
2. Newgard, C. B. & Attie, A. D. Getting biological about the genetics of diabetes. *Nature Med.* **16**, 388–391 (2010).
3. Illig, T. *et al.* A genome-wide perspective of genetic variation in human metabolism. *Nature Genet.* **42**, 137–141 (2010).
4. Gieger, C. *et al.* Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* **4**, e1000282 (2008).
5. Evans, A. M., DeHaven, C. D., Barrett, T., Mitchell, M. & Milgram, E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal. Chem.* **81**, 6656–6667 (2009).
6. Ohta, T. *et al.* Untargeted metabolomic profiling as an evaluative tool of fenofibrate-induced toxicology in Fischer 344 male rats. *Toxicol. Pathol.* **37**, 521–535 (2009).
7. Suhre, K. *et al.* Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS ONE* **5**, e13953 (2010).
8. Altmaier, E. *et al.* Bioinformatics analysis of targeted metabolomics—uncovering old and new tales of diabetic mice under medication. *Endocrinology* **149**, 3478–3489 (2008).
9. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
10. Chambers, J. C. *et al.* Genetic loci influencing kidney function and chronic kidney disease. *Nature Genet.* **42**, 373–375 (2010).
11. Köttgen, A. *et al.* New loci associated with kidney function and chronic kidney disease. *Nature Genet.* **42**, 376–384 (2010).
12. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature Genet.* **42**, 105–116 (2010).
13. Aulchenko, Y. S. *et al.* Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nature Genet.* **41**, 47–55 (2009).
14. Panneerselvam, K. & Freeze, H. H. Mannose enters mammalian cells using a specific transporter that is insensitive to glucose. *J. Biol. Chem.* **271**, 9417–9421 (1996).
15. Taguchi, T. *et al.* Hepatic glycogen breakdown is implicated in the maintenance of plasma mannose concentration. *Am. J. Physiol. Endocrinol. Metab.* **288**, E534–E540 (2005).
16. Blombaeck, B., Blombaeck, M., Edman, P. & Hessel, B. Amino-acid sequence and the occurrence of phosphorus in human fibrinopeptides. *Nature* **193**, 833–834 (1962).
17. Martin, S. C., Ekman, P., Forsberg, P. O. & Ersmark, H. Increased phosphate content of fibrinogen *in vivo* correlates with alteration in fibrinogen behaviour. *Thromb. Res.* **68**, 467–473 (1992).
18. Yuan, X. *et al.* Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. *Am. J. Hum. Genet.* **83**, 520–528 (2008).
19. Tregouet, D. A. *et al.* Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. *Blood* **113**, 5298–5303 (2009).
20. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
21. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genet.* **43**, 333–338 (2011).
22. Döring, A. *et al.* SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nature Genet.* **40**, 430–436 (2008).
23. Caulfield, M. J. *et al.* SLC2A9 is a high-capacity urate transporter in humans. *PLoS Med.* **5**, e197 (2008).
24. Klein, T. E. *et al.* Integrating genotype and phenotype information: an overview of the PharmGKB project. *Pharmacogenomics J.* **1**, 167–170 (2001).

25. Deeken, J. F. *et al.* A pharmacogenetic study of docetaxel and thalidomide in patients with castration-resistant prostate cancer using the DMET genotyping platform. *Pharmacogenomics J.* **10,** 191–199 (2010).
26. Lankisch, T. O. *et al.* Gilbert's Syndrome and irinotecan toxicity: combination with UDP-glucuronosyltransferase 1A7 variants increases risk. *Cancer Epidemiol. Biomarkers Prev.* **17,** 695–701 (2008).
27. Huang, R. S. *et al.* A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. *Proc. Natl Acad. Sci. USA* **104,** 9758–9763 (2007).
28. Chen, Y. *et al.* Effect of genetic variation in the organic cation transporter 2 on the renal elimination of metformin. *Pharmacogenet. Genomics* **19,** 497–504 (2009).
29. Shu, Y. *et al.* Effect of genetic variation in the organic cation transporter 1, OCT1, on metformin pharmacokinetics. *Clin. Pharmacol. Ther.* **83,** 273–280 (2008).
30. The SEARCH Collaborative Group. *SLCO1B1* variants and statin-induced myopathy—a genomewide study. *N. Engl. J. Med.* **359,** 789–799 (2008).
31. Davies, N. J. *et al.* AKR1C isoforms represent a novel cellular target for jasmonates alongside their mitochondrial-mediated effects. *Cancer Res.* **69,** 4769–4775 (2009).
32. Sanna, S. *et al.* Common variants in the *SLCO1B3* locus are associated with bilirubin levels and unconjugated hyperbilirubinemia. *Hum. Mol. Genet.* **18,** 2711–2718 (2009).
33. Johnson, A. D. *et al.* Genome-wide association meta-analysis for total serum bilirubin levels. *Hum. Mol. Genet.* **18,** 2700–2710 (2009).
34. Kolz, M. *et al.* Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet.* **5,** e1000504 (2009).
35. Zhai, G. *et al.* Eight common genetic variants associated with serum DHEAS levels suggest a key role in ageing mechanisms. *PLoS Genet.* **7,** e1002025 (2011).
36. Mootha, V. K. & Hirschhorn, J. N. Inborn variation in metabolism. *Nature Genet.* **42,** 97–98 (2010).
37. Meredith, D. & Christian, H. C. The SLC16 monocaboxylate transporter family. *Xenobiotica* **38,** 1072–1106 (2008).
38. Koepsell, H. & Endou, H. The SLC22 drug transporter family. *Pflugers Arch.* **447,** 666–676 (2004).

Author Contributions Designed the study: J.A., C.G., T.I., D.M., N.S. and K.S. Conducted the experiments: D.M., M.V.M. and R.P.M. Analysed the data: J.A., E.A., C.G., G.K., A.K., F.K., C.M., D.M., A.-K.P., C.P., J.R., J.S.R., W.R.-M., S.-Y.S., K.S. and B.W. Provided material, data or analysis tools: the CARDIoGRAM consortium, P.D., J.E., E.G., C.J.H., M.H.d.A., T.I., M.M., T.M., H.-W.M., N.J.S., K.S.S., T.D.S., H.-E.W. and G.Z. Wrote the paper: C.G., N.S. and K.S. All authors read the paper and contributed to its final form.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.S. (karsten@suhre.fr) or N.S. (ns6@sanger.ac.uk).

METHODS

Study populations. The KORA S4 survey, an independent population-based sample from the general population living in the region of Augsburg, southern Germany, was conducted in 1999−2001. The study design and standardized examinations of the survey (4,261 participants, response 67%) have been described in detail (ref 39 and references therein). A total of 3,080 subjects participated in a follow-up examination, KORA F4, in 2006−2008, comprising individuals who, at that time, were aged 32–81 years. The TwinsUK cohort is a British adult-twin registry in the age range 8–102 years and 84% are female. The samples used in this study are aged 23–85 (mean age 48 years) and 97% are female. These unselected twins were recruited from the general population through national media campaigns and were shown to be comparable to age-matched population singletons in terms of disease-related and lifestyle characteristics[40]. In both studies, written informed consent has been given by all participants and the studies have been approved by the local ethics committees (Bayerische Landesärztekammer for KORA and Guy's and St. Thomas' Hospital Ethics Committee for TwinsUK).

Blood sampling. Blood samples for metabolic analysis and DNA extraction from KORA were collected between 2006 and 2008 as part of the KORA F4 follow-up. To avoid variation due to circadian rhythm, blood was drawn in the morning between 08:00 and 10:30 after a period of at least 10 h overnight fasting. Material was drawn into serum gel tubes, gently inverted twice and then allowed to rest for 30 min at room temperature (18−25 °C) to obtain complete coagulation. The material was then centrifuged for 10 min (2,750g at 15 °C). Serum was divided into aliquots and kept for a maximum of 6 h at 4 °C, after which it was frozen at −80 °C until analysis. For the TwinsUK study, blood samples were taken after at least 6 h of fasting. The samples were immediately inverted three times, followed by 40 min of resting at 4 °C to obtain complete coagulation. The samples were then centrifuged for 10 min at 2,000g. Serum was removed from the centrifuged brown-topped tubes as the top, yellow, translucent layer of liquid. Four aliquots of 1.5 ml were placed into skirted microcentrifuge tubes and then stored at −45 °C until sampling.

Metabolomics measurements. Metabolon, a commercial supplier of metabolic analyses, developed a platform that integrates the chemical analysis, including identification and relative quantification, data-reduction and quality-assurance components of the process. The analytical platform incorporates two separate ultrahigh-performance liquid chromatography/tandem mass spectrometry (UHPLC/MS/MS2) injections and one gas chromatography/mass spectrometry (GC/MS) injection per sample. The UHPLC injections were optimized for basic and acidic species. A total of 295 metabolites were measured, spanning several relevant classes (amino acids, acylcarnitines, sphingomyelins, glycerophospholipids, carbohydrates, vitamins, lipids, nucleotides, peptides, xenobiotics and steroids; a full list of metabolites is given in Supplementary Table 1). The detection of the entire panel was carried out with 24 min of instrument analysis time (two injections at 12 min each), while maintaining low median process variability (<12% across all compounds). The resulting MS/MS^2 data were searched against a standard library generated by Metabolon that included retention time, molecular mass to charge ratio (m/z), preferred adducts and in-source fragments as well as their associated MS/MS spectra for all molecules in the library. The library allowed for the identification of the experimentally detected molecules on the basis of a multiparameter match without the need for additional analyses. Metabolon has shown in a recent publication that their integrated platform enabled the high-throughput collection and relative quantitative analysis of analytical data and identified a large number and broad spectrum of molecules with a high degree of confidence[5]. The Metabolon platform has, among other studies, been successfully applied in the analysis of the adult human plasma metabolome[41] and the identification of sarcosine as a biomarker for prostate cancer[42].

Quality control of metabolomics data. For this study we measured the Metabolon panel in human blood from 1,768 individuals of the KORA cohort and in 1,052 individuals of the TwinsUK cohort. Quality control data (relative standard deviation, upper and lower 95% confidence interval and minimum and maximum observed values in quality control samples) are reported in Supplementary Table 1. To avoid spurious false-positive associations due to small sample sizes, only metabolic traits with at least 300 non-missing values were included, and data points of metabolic traits that lay more than three standard deviations off the mean were excluded by setting them to 'missing' in the analysis: 276 of 295 available metabolites and 37,179 metabolite ratios satisfied this criterion in KORA, resulting in a total of 37,455 metabolic traits. For the TwinsUK study, identical selection criteria for metabolic traits were used, resulting in 258 metabolites and 32,499 metabolite ratios, and a total of 32,757 metabolic traits.

Genotyping and imputation. For all individuals profiled from the KORA study, genome-wide SNP data were already available. GWAS data of KORA and TwinsUK have been used and described extensively in the past, in the context of numerous GWAS and meta-analyses[3,34,43]. We therefore summarize only the essential details here. Genotyping of the KORA F4 population was carried out using the Affymetrix GeneChip array 6.0. Genotypes were determined using the Birdseed2 clustering algorithm. For quality assurance, we applied the criteria of call rate > 95% and P(Hardy-Weinberg) > 10^{-6} as filters for SNP quality: 655,658 autosomal SNPs satisfied these criteria. These genotyped SNPs were used for genome-wide analysis of the metabolic traits. For selection of the best-associated SNP within a region in a meta-analysis of KORA and TwinsUK, we used genotyped SNPs as well as dosages of imputed SNPs. In KORA F4, imputation was done using IMPUTE v0.4.2 (ref. 44) based on HapMap2 (see below).

Genotyping of the TwinsUK data set was done with a combination of Illumina arrays (HumanHap300, HumanHap610Q, 1M-Duo and 1.2MDuo 1M)[45,46]. We pooled the normalized intensity data for each of the three arrays separately (with 1M-Duo and 1.2MDuo 1M pooled together). For each data set, we used the Illluminus calling algorithm[47] to assign genotypes in the pooled data. No calls were assigned if an individual's most likely genotype was called with a posterior probability threshold of <0.95. Validation of pooling was achieved via a visual inspection of 100 random, shared SNPs for overt batch effects. Finally, intensity cluster plots of significant SNPs were visually inspected for over-dispersion-biased no calling, and/or erroneous genotype assignment. SNPs showing any of these characteristics were discarded.

We applied similar exclusion criteria to each of the three data sets separately. Exclusion criteria for samples were: (1) sample call rate <98%; (2) heterozygosity across all SNPs ≥ 2 s.d. from the sample mean; (3) evidence of non-European ancestry as assessed by principle-component-analysis comparison with HapMap3 populations; (4) observed pairwise identity-by-descent (IBD) probabilities indicative of sample identity errors. We corrected misclassified monozygotic and dizygotic twins on the basis of IBD probabilities. Exclusion criteria for SNPs were: (1) Hardy–Weinberg P value <10^{-6}, assessed in a set of unrelated samples; (2) minor allele frequency (MAF) <1%, assessed in a set of unrelated samples; (3) SNP call rate <97% (SNPs with MAF ≥ 5%) or <99% (for 1% ≤ MAF < 5%).

Alleles of all three data sets were aligned to HapMap2 or HapMap3 forward-strand alleles. Before merging, we performed pairwise comparison among the three data sets and further excluded SNPs and samples to avoid spurious geno-typing effects, identified as follows: (1) concordance at duplicate samples <1%; (2) concordance at duplicate SNPs <1%; (3) visual inspection of QQ-plots for logistic regression applied to all pairwise data-set comparisons; (4) Hardy–Weinberg P value <10^{-6}, assessed in a set of unrelated samples; (5) observed pairwise IBD probabilities indicative of sample identity errors. We then merged the three data sets, keeping individuals typed at the largest number of SNPs when an individual was typed at two different arrays. The merged data set consists of 5,654 individuals (2,040 from the HumanHap300, 3,461 from the HumanHap610Q and 153 from the HumanHap1M and 1.M arrays), and up to 874,733 SNPs depending on the data set (HumanHap300: 303,940, HumanHap610Q: 553,487, HumanHap1M and 1.M: 874,733). Imputation was performed using the IMPUTE software package (v2)[44], using two reference panels, P0 (HapMap2, rel 22, combined CEU+YRI+ASN panels) and P1 (610k+, including the combined HumanHap610k and 1M reduced to 610k SNP content). The analysis of this study used 534,665 autosomal SNPs (basically 610K SNPs extracted from the final merged data set).

Statistical analyses. The primary association testing was carried out using linear regressions on all metabolite concentrations and all possible ratios of metabolite concentrations. This was motivated by our previous observation[4,8] that the use of ratios may lead to a strong reduction in the overall trait variance. A test of normality showed that in 29,338 cases, the log-transformed ratio distribution was significantly better represented by a normal distribution than when untransformed ratios were used. In 5,145 cases, the untransformed distribution was closer to a normal distribution. For concentrations, 149 were closer to a log-normal distribution and 124 were better represented by a normal distribution. On the basis of this observation, and also for sake of simplicity, we decided to log-transform all metabolites and their ratios. We used P-gain statistics[4,8] to quantify the decrease in P value for the association with the ratio compared to the P values of the two corresponding concentrations. A high P-gain (more than 250) indicates that two metabolites are likely to be functionally linked in a metabolic pathway that has an impact on the associating genotype KORA and TwinsUK are population-based studies. They comprise only individuals who are not displaying any severe clinical symptoms at the time of sampling. Therefore, disease state was not considered as a confounding factor in the statistical analysis. In KORA, the software PLINK (version 1.06)[48] and SNPTEST was used with age and gender as covariates. To account for the family structure in the TwinsUK study, we used variance components applied to a score test implemented in the software Merlin[49].

Correction for multiple testing. We applied a conservative Bonferroni correction to control for false-positive error rates deriving from multiple testing. Using the KORA study as a reference, we corrected for tests on 655,658 SNPs and 37,455 metabolic traits, thus obtaining a Bonferroni-adjusted P value of $P = 2.04 \times 10^{-12}$. For ratios, we also required that the increase in the strength of association, expressed as the change in P value when using ratios compared to the larger of the two P values when using two metabolite concentrations individually (P-gain), should be larger than the number of tested metabolic traits (P-gain > 250)[4,8]. This limit is considered a Bonferroni-type conservative cutoff for identifying those metabolite concentration pairs for which the use of ratios strongly improves the strength of association. In addition to the strongest associating metabolic trait, others can often provide additional insight into the underlying biochemical processes. In such cases, we consider a P value of 1.33×10^{-6} to represent a conservative level of significance (Bonferroni correction for 37,455 tests at a nominal significance level of 5%).

Inflation. In most cases, the assumption of a linear additive model was valid (see box plots in Supplementary Fig. 3) and there was no inflation of summary statistics, which could be indicative of population stratification (see QQ-plots in Supplementary Fig. 3). Lambda values ranged from 0.965 to 1.024 (median = 1.006) in KORA, and from 0.940 to 1.013 (median = 0.985) in TwinsUK.

Candidate gene selection and overlap with disease loci. Regional association plots (Supplementary Fig. 3) were created using imputed and meta-analysed data. Within this region, the SNP with the strongest signal of association in the meta-analysis was retained as the final SNP to be reported. Association data for all metabolic traits at the 37 SNPs reported in Table 1 (for KORA, TwinsUK and meta-analysis), limited to associations with $P < 1.33 \times 10^{-6}$ (Bonferroni correction for multiple testing of metabolic traits at a single locus) and P-gain > 250 (for ratios) in the meta-analysis, are reported in Supplementary Table 4. For the strongest associating trait, box plots were plotted to visualize the actual quantitative dependence of the trait on genotype (Supplementary Fig. 3). On the basis of association data alone, it is not possible in most cases to identify the gene within a locus that causes the association. However, using knowledge of the function of genes in LD with the reported SNP, as well as the biochemical characteristics of the associating metabolite, it is possible to identify a single most likely candidate gene in many cases. These cases are tagged as 'match between gene function and metabolic trait' and are supported by arguments provided as Supplementary Text (for example, the association between a SNP in LD with *OPLAH* (oxoprolinase) and oxoproline concentrations). At two loci (*CYP4A* and *UGT1A*), alternative splice variants exist. We named these loci without attempting to specify the exact variant.

GWAS catalogue. Using the catalogue of published GWAS (accessed 10 October 2010)[1], we identified for each entry the SNPs in the KORA and TwinsUK studies that correlate most strongly with a previously reported SNP ($r^2 \geq 0.5$) and that were present in our association database ($P < 10^{-3}$, P-gain > 10). The resulting associations are available online on our GWAS server. New associations will be included as the database of published GWAS is updated.

Enrichment analysis. We downloaded the actual version of the GWAS catalogue from NHGRI and deleted all records that correspond to our previous studies. As a sampling data set, we chose the 655,658 SNPs from the Affymetrix 6.0 array, which have been tested in the KORA part of this study. The 37 SNPs that we report are from this array and can therefore be considered to represent one draw-out of this set. We then drew 1,000,000 sets of 37 SNPs at random (with replacement) from this sampling data set. To account for comparable MAF distributions between the reference and the random set, we then rejected all draws in which the mean or the variance of the MAF distributions were significantly different ($P < 0.05$) between the random and the reference set: 330,775 random sets were hence retained. Using an LD criterion of $r^2 > 0.8$ (based on HapMap2 release 27, NCBI B36, CEU population), we then counted the overlap with the GWAS catalogue for every random set. The reference set was included as a technical positive control in the computations. For the 330,775 tested random sets, at most six overlapping SNPs were found (8 times), and in more than half of the cases, no overlapping SNPs were present in the sampled data set (see Supplementary Table 9). For our reported 37 metabolomics SNPs, we identified 14 overlapping SNPs (note that we report 15 overlapping loci in Fig. 1; the *ENPEP* locus was not yet included in the GWAS catalogue and was not used in this analysis). Because we never found 14 overlapping loci by chance, the P value of our observations being due to chance is less than $1/330{,}775 \approx 3 \times 10^{-6}$.

Functional characterization of SLC16A9. The *SLC16A9* (*MCT9*) clone (IMAGE ID 40146598) was purchased from Autogen Bioclear. Plasmid was linearized with SpeI restriction enzyme (New England Biolabs) and complementary RNA was synthesized *in vitro* using the T7 mMachine *in vitro* transcription system (Ambion). *MCT9* was expressed in *Xenopus laevis* oocytes as described previously[50]. Briefly, oocytes at stage V–VI were injected with 10 ng of *MCT9* cRNA and incubated in modified Barth's solution for 3–4 days at 18 °C with the medium changed daily. Control oocytes had either no injection or an injection of an equal volume (50 nl) of distilled water, and were incubated for the same length of time. Uptake and efflux experiments were performed similarly to those described previously[50] except that the substrate was [^{3}H]-carnitine (specific activity 81 Ci mmol^{-1}, GE Healthcare).

Data access. This study generated millions of individual data points through the profiling of n metabolites ($n = 250$) and $n(n-1)/2$ ratios in about 2,800 individuals, and the subsequent associations with millions of genetic variants from GWAS. We created a web-based interface and visualization tools for the dissemination of results to the scientific community, with the aim of allowing rapid storage and retrieval of data as well as managing the integration of metabolomics summary statistics vis-à-vis published GWAS studies. The association data are freely available at http://www.gwas.eu and at mirror sites located at the Wellcome Trust Sanger Institute and King's College London sites.

Web links. GWAS server: http://www.gwas.eu, SNAP: http://www.broadinstitute.org/mpg/snap/, NHGRI catalogue of published GWAS: http://www.genome.gov/gwastudies/, eQTL: http://www.sanger.ac.uk/Software/analysis/genevar/, GRAIL: http://www.broadinstitute.org/mpg/grail/, IPA (Ingenuity Pathway Analysis): http://ingenuity.com, OMIM: http://www.ncbi.nlm.nih.gov/omim, yED network editor: http://www.yworks.com, BioGPS: http://biogps.gnf.org, Genecards: http://www.genecards.org, WikiGenes: http://www.wikigenes.org, Pharmacogenomics knowledge base: http://www.pharmgkb.org, R statistical analysis system: http://www.r-project.org, KORA study population: http://www.helmholtz-muenchen.de/kora/, TwinsUK study: http://www.twinsuk.ac.uk, Metabolon Inc.: http://www.metabolon.com, MERLIN: http://www.sph.umich.edu/csg/abecasis/Merlin, PLINK: http://pngu.mgh.harvard.edu/~purcell/plink, R: http://www.r-project.org, SNPTEST: http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest.html

39. Wichmann, H. E., Gieger, C. & Illig, T. KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67** (Suppl 1), 26–30 (2005).
40. Andrew, T. *et al.* Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res.* **4**, 464–477 (2001).
41. Lawton, K. A. *et al.* Analysis of the adult human plasma metabolome. *Pharmacogenomics* **9**, 383–397 (2008).
42. Sreekumar, A. *et al.* Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **457**, 910–914 (2009).
43. Soranzo, N. *et al.* A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nature Genet.* **41**, 1182–1190 (2009).
44. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
45. Richards, J. B. *et al.* Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* **371**, 1505–1512 (2008).
46. Soranzo, N. *et al.* Meta-analysis of genome-wide scans for human adult stature identifies novel loci and associations with measures of skeletal frame size. *PLoS Genet.* **5**, e1000445 (2009).
47. Teo, Y. Y. *et al.* A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* **23**, 2741–2746 (2007).
48. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
49. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genet.* **30**, 97–101 (2002).
50. Meredith, D. Site-directed mutation of arginine 282 to glutamate uncouples the movement of peptides and protons by the rabbit proton-peptide cotransporter PepT1. *J. Biol. Chem.* **279**, 15795–15798 (2004).

The mechanism of membrane–associated steps in tail–anchored protein insertion

Malaiyalam Mariappan[1]*, Agnieszka Mateja[2]*, Malgorzata Dobosz[2], Elia Bove[2], Ramanujan S. Hegde[1]† & Robert J. Keenan[2]

Tail–anchored (TA) membrane proteins destined for the endoplasmic reticulum are chaperoned by cytosolic targeting factors that deliver them to a membrane receptor for insertion. Although a basic framework for TA protein recognition is now emerging, the decisive targeting and membrane insertion steps are not understood. Here we reconstitute the TA protein insertion cycle with purified components, present crystal structures of key complexes between these components and perform mutational analyses based on the structures. We show that a committed targeting complex, formed by a TA protein bound to the chaperone ATPase Get3, is initially recruited to the membrane through an interaction with Get2. Once the targeting complex has been recruited, Get1 interacts with Get3 to drive TA protein release in an ATPase–dependent reaction. After releasing its TA protein cargo, the now–vacant Get3 recycles back to the cytosol concomitant with ATP binding. This work provides a detailed structural and mechanistic framework for the minimal TA protein insertion cycle.

Approximately 5% of eukaryotic membrane proteins are anchored to the lipid bilayer by a single carboxy-terminal transmembrane domain[1–4] (TMD). These 'tail-anchored' proteins are found in virtually all cellular membranes and perform essential functions in processes including protein trafficking, degradation, cell death and membrane biogenesis. TA proteins in compartments of the secretory and endocytic pathways are first targeted to and inserted into the ER membrane by a post-translational targeting pathway conserved across eukaryotes[5–9] and archaea[10,11].

This pathway begins with a 'pre-targeting' factor that captures newly synthesized TA proteins through their TMDs near the ribosome[12,13]. In yeast, the pre-targeting factor is Sgt2, which assembles with Get3, Get4 and Get5 (also known as Mdy2) to form a TMD recognition complex[12,14,15]. Assembly of TMD recognition complexes permits substrates to be transferred from Sgt2 to Get3 in an ATP-dependent manner[12]. Get3 (TRC40, or ASNA1, in mammals) is a homodimeric ATPase whose conformation is regulated by its nucleotide state[16–20]. Both crystallographic and functional analyses support a model in which an ATP-bound, 'closed' dimer of Get3 binds substrates in a large hydrophobic groove that spans both subunits[16–18,20]. This substrate–Get3–nucleotide complex is therefore the committed targeting complex (Supplementary Discussion).

In yeast, genetic and physical interaction studies have identified the ER-localized membrane proteins Get1 and Get2 as potential receptors for Get3 (refs 7, 21). It is not known whether Get1, Get2 and Get3 constitute the minimal targeting and insertion machinery, how they function or what their essential roles are during TA protein insertion. In this Article, we combine functional reconstitution of TA protein insertion with structural analysis of key intermediate complexes to provide a mechanistic framework for the TA protein insertion cycle in *Saccharomyces cerevisiae*.

The minimal insertion machinery

We first reconstituted the TA protein insertion cycle with purified recombinant factors. A functional TA protein targeting complex was assembled and purified from *in vitro* translation reactions (Supplementary Fig. 1). The complex contained radio-labelled and epitope-tagged Sec61β (an ER-localized TA protein) bound to recombinant yeast Get3 in roughly the 2:1 ratio expected from structural studies. This recombinant targeting complex was functional as judged by membrane insertion of Sec61β into ER-derived yeast rough microsomes (yRMs) but not into protein-free liposomes (Fig. 1a). Microsomes from ΔGet1 and ΔGet2 yeast strains showed little insertion activity, whereas ΔGet3 microsomes were similar to wild-type yRMs. Sec61β insertion efficiency with the purified targeting complex was approximately two-fold higher than for Sec61β in crude translation reactions (data not shown), consistent with the observation that the latter contains a heterogeneous mixture of Sec61β complexes with other factors[8,13,22]. Thus, purified Get3–Sec61β is a committed targeting complex for Get1- and Get2-dependent membrane insertion.

The TA insertion defect of ΔGet1 and ΔGet2 microsomes is due solely to loss of Get1 and/or Get2. To show this, purified recombinant Get1 and Get2 (rGet1 and rGet2; Supplementary Fig. 2) produced from *Escherichia coli* were added to detergent extracts prepared from ΔGet1 or ΔGet2 yRMs, reconstituted into proteoliposomes and tested for function (Supplementary Fig. 3). Proteoliposomes from ΔGet1 yRMs were inactive for TA protein insertion, but were restored by replenishment with physiologic levels of rGet1 but not rGet2. ΔGet2 proteoliposomes required both rGet1 and rGet2 to restore insertion to near wild-type levels (Supplementary Fig. 3), as expected because Get1 is absent from ΔGet2 yRMs (Fig. 1a). We also biochemically depleted Get1 and Get2 from wild-type yRM and showed that the resulting insertion defect could be corrected by replenishment with rGet1 and rGet2 but with neither individually (Supplementary Fig. 4). Thus, rGet1 and rGet2 are fully functional in replacing their native counterparts during Get3-dependent TA protein insertion.

The lack of membrane proteins co-purifying with Get1 and Get2 (Supplementary Fig. 5), and the absence of other membrane proteins found in genetic studies[15,23,24], suggested that Get1 and Get2 are

[1]Cell Biology and Metabolism Program, National Institute of Child Health and Human Development, National Institutes of Health, Room 101, Building 18T, 18 Library Drive, Bethesda, Maryland 20892, USA. [2]Department of Biochemistry & Molecular Biology, The University of Chicago, Gordon Center for Integrative Science, Room W238, 929 East 57th Street, Chicago, Illinois 60637, USA. †Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK.
*These authors contributed equally to this work.
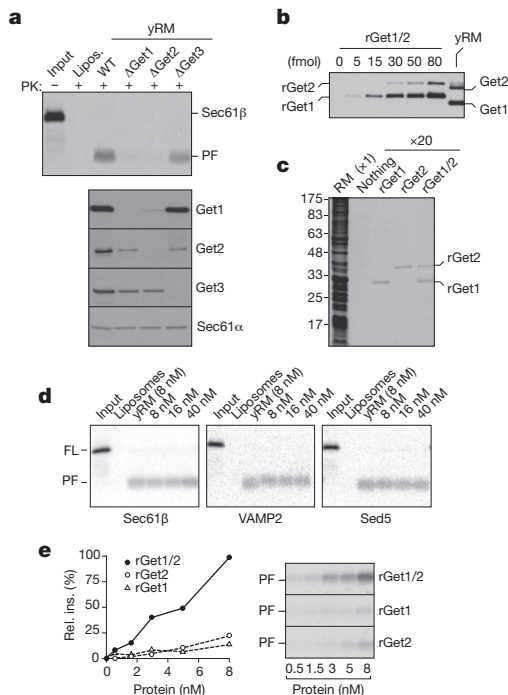
Figure 1 | Reconstitution of TA protein insertion with purified components. a, Yeast rough microsomes (yRMs) from the indicated strains were tested for insertion of purified Get3–Sec61β targeting complex (top) or by immunoblotting (bottom). The protease-protected fragment (PF) is diagnostic of successful insertion. Liposomes are a negative control. PK, proteinase K; WT, wild type. **b,** Quantification of Get1 and Get2 concentrations in yRMs by immunoblotting. **c,** Protein composition of yRMs and proteoliposomes containing recombinant proteins. Proteoliposomes in 20-fold relative excess were analysed. **d,** Insertion of purified targeting complexes into liposomes, yRMs, or rGet1/2 proteoliposomes. VAMP2 and Sed5, TMDs from rat VAMP2 or yeast Sed5. Concentrations of the Get1/2 complex are indicated FL, full length. **e,** Relative efficiency of insertion of purified Get3–Sec61β targeting complex into rGet1, rGet2 or rGet1/2 proteoliposomes. Autoradiographs and quantified data are shown.

sufficient for Get3-mediated TA protein insertion. Indeed, proteoliposomes containing physiologic concentrations of only rGet1 and rGet2 (Fig. 1b, c) were indistinguishable from yRM in mediating insertion of three different purified TA protein targeting complexes (Fig. 1d). Incorporating super-physiologic levels of rGet1 and rGet2 did not further improve insertion (Fig. 1d), and lower levels reduced overall insertion efficiency (Fig. 1e).

The recombinant system required both rGet1 and rGet2 (Fig. 1e), precisely mirroring the results *in vivo*[7] and in crude proteoliposomes (Supplementary Figs 3 and 4). Interaction analysis confirmed that rGet1 and rGet2 form a complex through their membrane domains in detergent solution (Supplementary Fig. 6), suggesting that during reconstitution they are incorporated as a complex. Taken together, the dependence on rGet1 and rGet2, their interaction with each other, their functionality in replacing the endogenous proteins and the high-efficiency insertion at native concentrations argue strongly that we have reconstituted physiologically relevant TA protein insertion with a defined targeting complex and only two membrane proteins.

The Get2c–Get3–ADP•AlF$_4^-$ complex

Membrane targeting presumably involves an interaction between Get3 and the conserved cytosolic domains of Get1 and/or Get2 (Figs 2a and 3a and Supplementary Fig. 10). These fragments ('Get1c' and 'Get2c') did not interact with each other (Supplementary Figs 6 and 7), but both bound tightly to Get3 (Supplementary Figs 7 and 8) and inhibited the insertion of Sec61β into yRMs (Supplementary Fig. 8). The ability of Get3 to interact with either subunit of the Get1/2 complex suggested that each interaction might serve a different purpose in the insertion cycle.

The closed-dimer form of ADP•AlF$_4^-$-bound Get3 probably mimics the TA substrate-bound conformation that targets to the membrane[16–18,20]. This Get3–ADP•AlF$_4^-$ complex crystallized with Get2c, and we determined the structure to a resolution of 2.1 Å (Supplementary Table 1 and Supplementary Fig. 9). The structure reveals Get3 in a 'closed'-dimer conformation with ADP•AlF$_4^-$ bound at each active site (Fig. 2b). Two Get2 fragments, each comprising two helices connected by a short linker, bind to equivalent sites on opposite faces of the symmetric Get3 homodimer. Each interface buries ~960 Å^2 of surface area, largely restricted to a single Get3 monomer (Fig. 2c, green, and
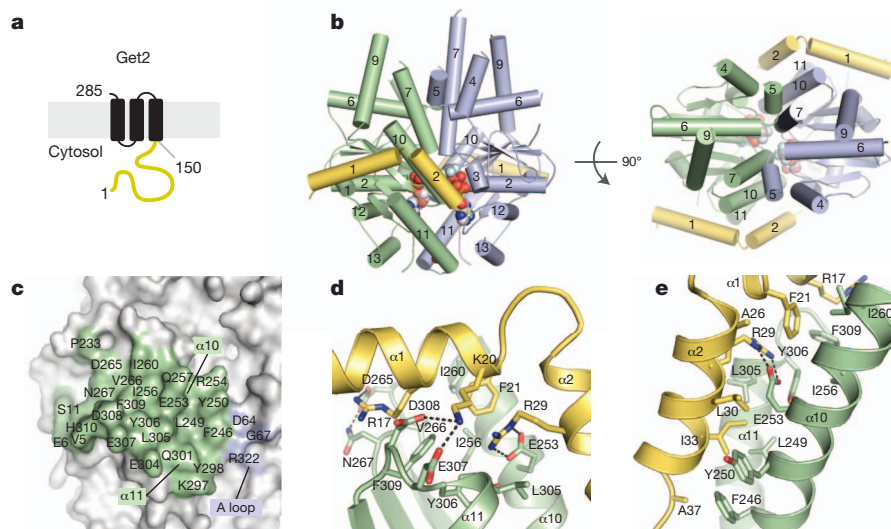
Figure 2 | Get2 fragment complex with ADP•AlF$_4^-$-bound Get3.
a, Predicted topology of *S. cerevisiae* Get2 with its large cytosolic-facing region (yellow). **b,** Structure of two Get2 fragments (yellow) bound to the closed Get3 dimer (green, blue). Two Mg^{2+}-ADP•AlF$_4^-$ complexes and a zinc atom are indicated (spheres). An orthogonal view into the substrate-binding composite hydrophobic groove is shown on the right. **c,** Get3 residues in the Get2 interface

are indicated. Most contacts are to one Get3 monomer (green); poorly ordered contacts are to the conserved A-loop ATPase motif. **d,** Close-up of interactions along helix α1 of Get2, including Arg 17, Lys 20 and Phe 21. **e,** Close-up of interactions along helix α2 of Get2, including the conserved salt bridge between Arg 29 and Glu 253.

Supplementary Fig. 10). Get3 residues within the interface undergo little conformational change on binding to Get2c (Supplementary Fig. 11). The amino-terminal helix of Get2 lies in a cleft defined at one end by short loops following helices α10 and α11 of Get3, and at the other end by the loop following helix α9 and the extreme N terminus of Get3 (Fig. 2d). Three conserved, negatively charged residues in Get3, namely Asp 265, Glu 307 and Asp 308, make direct contact with Get2c. The second helix of Get2 lies in a cleft defined by Get3 helices α10 and α11 (Fig. 2e). This surface is largely hydrophobic except for a conserved salt bridge between Glu 253 (Get3) and Arg 29 (Get2c). The C-terminal end of the Get2 fragment, which is not conserved, makes poorly ordered contacts with the adjacent Get3 monomer (Fig. 2c, blue).

The TA substrate-binding site in Get3 comprises a large hydrophobic groove spanning the α-helical subdomains of both monomers[16]. In the Get2–Get3 complex, this groove is intact (Fig. 2b and Supplementary Fig. 20), suggesting that Get2 captures the closed Get3 targeting complex without disrupting the TA binding site. The long, flexible linker that tethers the helical N terminus of Get2 to its first TMD would facilitate this process. Thus, we propose that the Get2c–Get3– ADP•AlF$_4^-$ structure represents a snapshot of the initial encounter between the closed-dimer targeting complex and the receptor.

The Get1c–Get3 complex

Get3 was also crystallized in the presence of Get1c. Whether or not ADP•AlF$_4^-$ was present during crystallization, the Get3–Get1c crystals lacked nucleotide. We determined the structure of this nucleotide-free complex to a resolution of 3.0 Å (Supplementary Table 1 and Supplementary Fig. 9) and revealed Get3 in an 'open' conformation, with two Get1 fragments bound to equivalent sites on opposite faces of the symmetric Get3 homodimer (Fig. 3b). Each Get1 fragment adopts an antiparallel coiled-coil structure and buries ~1,030 Å^2 of surface area in a bipartite interface split evenly between the two Get3 subunits (Fig. 3c and Supplementary Fig. 10). As observed in the Get2c complex, Get3 residues on the interface undergo little conformational change on binding to Get1c (Supplementary Fig. 11). Binding to one Get3 monomer is primarily mediated by hydrophobic

contacts between helix α2 of Get1c and the cleft defined by helices α10 and α11 of Get3 (Fig. 3c, d, green). Binding to the other monomer is mediated by helix α1 of Get1c, which interacts with Get3 helices α4 and α5, and by a six-residue loop in Get1c that directly contacts the ATP-binding site (Fig. 3c, e, blue; see below).

Importantly, many of the Get3 residues that contact Get1c also mediate interactions with Get2c (Supplementary Figs 10 and 11). For example, the conserved Arg 73 (Get1c)/Glu 253 (Get3) salt bridge almost perfectly mimics the Arg 29 (Get2c)/Glu 253 (Get3) inter-action (Figs 2e and 3d). The presence of overlapping binding sites suggests that Get1 and Get2 cannot simultaneously occupy the same site on Get3, as illustrated by dissociation of the Get3–Get2c complex by Get1c (Supplementary Fig. 11). Previous work underscores the functional significance of this region of Get3: alanine substitutions within the shared interface, including F246A, Y250A, E253A and Y298A, have a strong loss-of-function phenotype in yeast[18]. Moreover, two of these positions, Tyr 250 and Glu 253, have been implicated in the ATP-dependent binding of Get4[25]. Thus, the α10– α11 region of Get3 is a binding hotspot that probably plays an import-ant regulatory role at different stages of the targeting cycle.

The most striking aspect of the Get3–Get1c structure is how the Get1 coiled coil wedges between the Get3 subunits, completely dis-rupting the hydrophobic TA substrate-binding site (Fig. 3b). Such an interaction could effect substrate release from the Get3 targeting com-plex. However, parts of the bipartite Get1-binding site on Get3— including the ATPase motifs and portions of helices α4 and α5 (Fig. 3c, blue)—are buried in the ATP-bound, fully closed-dimer conformation. By contrast, the bipartite Get1-binding site is largely exposed to solvent in the Mg^{2+}-ADP-bound state[17,20] (Supplementary Fig. 12). This implies that ATP hydrolysis by the targeting complex is needed to expose the Get1-binding site on Get3 (Fig. 3c and Sup-plementary Fig. 12, green and blue). Once exposed, Get1 would complete the Get3 transition from closed to open, disrupting the hydrophobic groove to promote release of the TA substrate and ADP (which binds weakly to substrate-free Get3; Supplementary Fig. 18). Importantly, the rigid Get1 coiled coil is perpendicular to
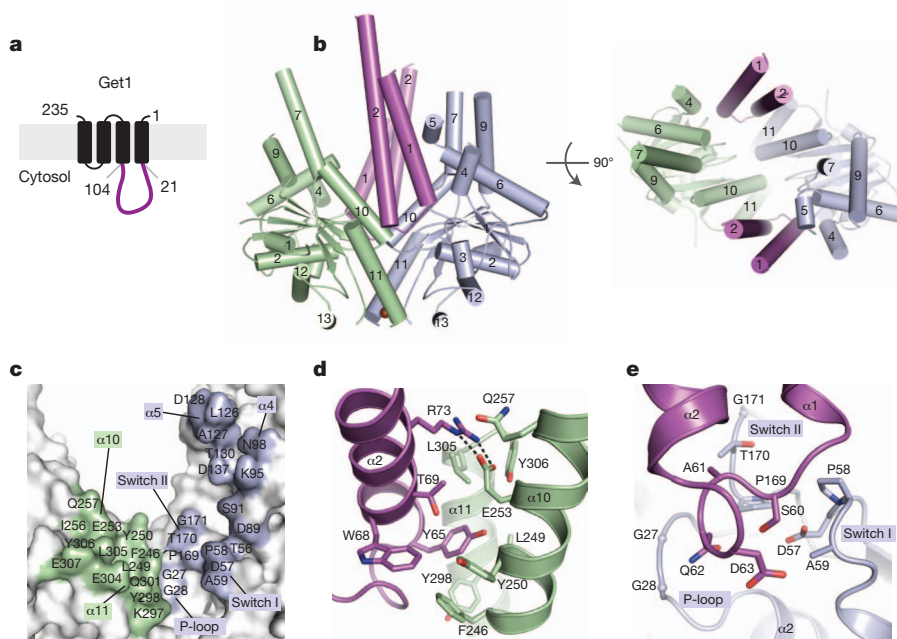
Figure 3 | Get1 fragment complex with Get3. a, Predicted topology of *S. cerevisiae* Get1 with a large cytosolic-facing region (magenta). **b,** Structure of two Get1 fragments (magenta) bound to the open dimer state of Get3 (green, blue). The composite hydrophobic groove is completely disrupted. **c,** Get3 residues in the Get1 interface are indicated; significant contacts are made to both monomers (green, blue), including the P-loop, switch I and switch II

ATPase motifs. **d,** Close-up of interactions between Get1 helix α2 (magenta) and one Get3 monomer (green), including the conserved salt bridge between Arg 73 and Glu 253. This interface overlaps extensively with the Get2c binding surface (Fig. 2e and Supplementary Fig. 11). **e,** Close-up of interactions between the Get1 hairpin loop and the active site of the adjacent Get3 monomer (blue).

the plane of the membrane, thereby positioning the hydrophobic groove of Get3 parallel to the membrane. This implies that the TMD of a TA protein is precisely released along the membrane surface, presumably facilitating its subsequent insertion.

Targeting and substrate release

Conserved contacts between Get3–Get2 and Get3–Get1 were disrupted with point mutations (R17E and R73E, respectively), verified to prevent binding (Supplementary Fig. 13) and shown to reduce insertion in the reconstituted system sharply (Fig. 4a). When the substrate–Get3 interaction was monitored by crosslinking (Supplementary Fig. 14), Get1c, but not Get2c, was found to release TA substrate from Get3 (>50% at 500 nM; Fig. 4b). This activity was abolished by the R73E mutation that disrupts Get3–Get1c interactions (Fig. 4c). Thus, Get1c and Get2c both inhibit insertion (Supplementary Fig. 8), but for different reasons: Get1c causes premature substrate release whereas Get2c competitively precludes targeting.

When reconstituted into proteoliposomes at more-physiologic concentrations, neither rGet1 nor rGet2 was able to effect substrate release, whereas the complete rGet1/2 complex was active (Fig. 4d). Importantly, disrupting the Get3–Get1 interaction (with R73E) or the Get3–Get2 interaction (with R17E) abolished the ability of the rGet1/2 complex to stimulate substrate release (Fig. 4e). Thus, whereas Get1c at super-physiologic concentrations can drive substrate release on its own, full-length Get1 in the membrane is unable to do so at physiologic levels. In this context, Get1 requires Get2 (specifically its ability to bind Get3) to release substrate from Get3.

On the basis of the Get3–Get1c structure, ATP hydrolysis by the Get3 targeting complex is likely to be necessary for its interaction with Get1. Indeed, targeting complexes containing an ATPase-deficient Get3 mutant (D57N) were poorly inserted into proteoliposomes containing the rGet1/2 complex (Fig. 4f) despite no impairment of the interaction of Get3 (D57N) with substrate or the rGet1/2 complex (Supplementary Fig. 15 and data not shown). Analysis of the interaction between TA substrate and Get3 (D57N) revealed that the rGet1/2 complex was unable to induce release (Fig. 4d, e). Taken together, the results of the functional analysis indicate that the Get3–Get2 interaction is important for targeting, and that this step

is a prerequisite for substrate release. Substrate release, in turn, depends on both ATP hydrolysis by Get3 and the ability of Get3 to interact with Get1.

ATP–dependent recycling

The ATP that Get3 hydrolyses before substrate release is apparently acquired from the *in vitro* translation reaction (and maintained during purification) because insertion proceeds efficiently without additional ATP in the purified system (Fig. 5a). This is consistent with structural analysis suggesting that nucleotide is shielded from bulk solvent in the fully closed Get3–ATP–substrate ternary complex (Supplementary Discussion). However, we (Supplementary Fig. 16) and others[6,17] have found that insertion reactions into crude yRMs, but not rGet1/2 proteoliposomes, are stimulated by ATP, non-hydrolysable ATP analogues or ADP. The explanation for this discrepancy proved to be the near-stoichiometric presence of Get3 on the Get1/2 complex in yRMs (Supplementary Fig. 5), but not on rGet1/2 proteoliposomes. Accordingly, binding Get3 to rGet1/2 proteoliposomes restored ATP dependence (Fig. 5b), whereas removing Get3 from yRM (by using ΔGet3 yeast) eliminated the ATP requirement for maximal insertion (Fig. 5c).

These results indicate that after TA substrate release, Get3 remains bound to microsomal membranes. In the nucleotide-free Get3–Get1c structure, which mimics this 'post-insertion' complex, residues within the conserved loop of Get1 ([59]ISAQDN[64]) insert into the Get3 active site (Fig. 3e) and deform it relative to the ADP•AlF$_4^-$-bound conformation (Fig. 5d). Modelling ATP into the active site reveals steric and electrostatic clashes between Get1 and ATP, suggesting that free ATP should displace Get3 from Get1. Indeed, the Get3–Get1c interaction was quantitatively disrupted by micromolar concentrations of ATP (Fig. 5e). ADP was far less effective, and AMP failed to disrupt the Get3–Get1c complex. This ATP-dependent Get3 dissociation was also verified with full-length Get1 using pull-down assays (Supplementary Fig. 19). By contrast, none of the tested nucleotides

Figure 4 | Mutational analysis of the function of Get1, Get2 and Get3. a, Insertion assay with purified Get3–Sec61β targeting complex and proteoliposomes containing the indicated purified proteins. Liposomes and yRM are controls. Get1* and Get2* indicate mutants inactive in Get3 interaction (R73E and R17E, respectively). **b**, Substrate release from targeting complexes incubated with Get1c or Get2c; release was monitored by loss of the crosslink (XL) between radio-labelled substrate and Get3. Square brackets indicate concentration. **c**, As in **b**, with wild-type and mutant fragments at a concentration of 0.5 μM. **d**, Substrate interaction with Get3 or the ATPase-deficient Get3 (D57N) was assessed by crosslinking after incubation with liposomes or proteoliposomes containing the indicated recombinant proteins. **e**, As in **d**, but comparing wild-type and mutant complexes of Get1 and Get2. **f**, Relative efficiency of insertion (mean ± s.e.m.; n = 6) into rGet1/2 proteoliposomes with targeting complexes prepared from wild-type Get3 or Get3 (D57N).

Figure 5 | ATP-dependent recycling of empty Get3 from Get1. a, Insertion activity of purified Get3–Sec61β targeting complex using the indicated vesicles with or without an ATP regenerating system. **b**, Proteoliposomes containing the rGet1/2 complex, or the rGet1/2 complex bound to Get3 (left panel), were tested for insertion activity of purified targeting complex in the presence or absence of ATP (right panel). Coom., Coomassie blue. **c**, Purified targeting complex was tested for insertion into wild-type yRMs or those from a ΔGet3 strain, with or without ATP. **d**, Close-up of the Get1c–Get3 complex (magenta and blue) modelled onto the active site of the closed, ADP•AlF$_4^-$-bound Get3 dimer (grey). Steric (dashed lines) and electrostatic clashes between conserved residues in Get1 and the nucleotide γ-phosphate are apparent. **e**, Dissociation of Get3–Get1c, monitored by the change in fluorescence resonance energy transfer (ΔF), on titration with the indicated nucleotides. Curve fits of triplicate measurements (mean ± s.e.m.) are shown. a.u., arbitrary units. The reaction contained 10 nM Get3 (D57N) and 100 nM Get1c. **f**, As in **e**, but with 10 nM Get3 (D57N) and 200 nM Get2c.
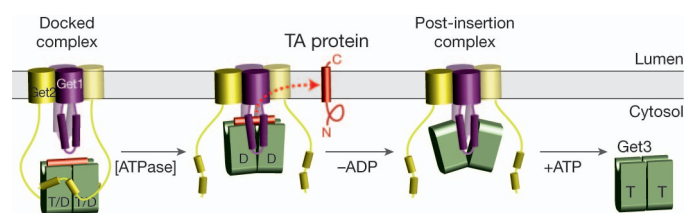
Figure 6 | Model for TA protein insertion. Nucleotide- and tail-anchored substrate-bound Get3 in a closed-dimer conformation forms the 'docked complex' by association with Get2. D, ADP; T, ATP. Following ATP hydrolysis, Get1 interacts with and orients Get3 along the membrane surface. This stabilizes the open-dimer conformation of Get3, disrupts the composite hydrophobic groove and promotes TA substrate release for membrane insertion. The Get3–Get1 post-insertion complex is dissociated by ATP binding, recycling Get3 back to the cytosol. See Supplementary Discussion for more details.

disrupted Get2c binding to Get3 (Fig. 5f). Thus, free ATP binding dissociates the Get1–Get3 complex to recycle Get3 from the membrane after TA substrate release.

A model for the insertion cycle

Figure 6 illustrates our working framework for the insertion cycle. Substrate-bound Get3 in the closed conformation and loaded with nucleotide (either ATP or ADP; see Supplementary Discussion) is captured at the membrane by the cytosolic domain of Get2. The apparently long and flexible Get2 tether may facilitate this initial encounter and bring the intact targeting complex near to the site of insertion. After this targeting step, Get1 mediates the post-targeting reactions of substrate release and insertion. Get1 binding to the targeting complex would be facilitated by partial destabilization of the closed dimer after ATP hydrolysis, and by the high local concentration of Get3 achieved by its recruitment through Get2. Binding to the rigid Get1 coiled coil would orient Get3 such that the substrate is in close proximity to the membrane. Moreover, by stabilizing the open conformation, Get1 binding would disrupt the Get3 hydrophobic groove and promote release of substrate and ADP. At present, we do not know whether the Get1/2 complex functions as a heterodimer or heterotetramer, although we favour the latter given the symmetric structure of the Get3 dimer. The released substrate would insert unassisted into the lipid bilayer directly[26,27] or would be chaperoned by the TMDs of the Get1/2 complex. Finally, the empty Get3 would be released from Get1 concomitant with ATP binding, and would be primed to accept the next substrate from the cytosolic pre-targeting complex for another round of targeting.

METHODS SUMMARY

Reagents and assays. Constructs, proteins and antibodies derived from earlier studies[8,13,16] are described in Methods. Antibodies against Get1 and Get2 were produced in rabbits. *In vitro* translation, insertion, crosslinking and immunoprecipitation were as described previously[8,13,28]. Get1 and Get2 (full length and fragments) were expressed in *E. coli* and purified by Ni-NTA chromatography; fragments were further purified by size exclusion chromatography. ^{35}S-labelled targeting complexes were affinity-purified from *in vitro* translation reactions containing recombinant Get3.
Liposomes, microsomes and proteoliposomes. Liposomes containing a 4:1 ratio of egg phosphatidylcholine and dipalmitoylphosphatidylethanolamine were prepared by extrusion[27,29]. Yeast rough microsomes were prepared as before[30–33]. Proteoliposome reconstitutions from solubilized yRMs or purified Get1 and/or Get2 were done by optimizing (Supplementary Fig. 17) earlier methods[30,33,34].
Interaction analysis. Binding assays were performed by gel filtration and multi-angle light scattering, pull-down assays or fluorescence resonance energy transfer. Substrate release was monitored by amine-reactive crosslinking[8].
Structure determination. Complexes of Get3 with Get1c or Get2c were co-expressed in *E. coli* and purified by Ni-NTA and size exclusion chromatography. Diffraction data were collected at Advanced Photon Source beamline 21-IDG, Argonne National Laboratory. Structures were determined by molecular replacement in PHASER[35]. Refinement and model building was done using PHENIX[36] and COOT[37].

1. Kutay, U., Hartmann, E. & Rapoport, T. A. A class of membrane proteins with a C-terminal anchor. *Trends Cell Biol.* **3**, 72–75 (1993).
2. Beilharz, T., Egan, B., Silver, P. A., Hofmann, K. & Lithgow, T. Bipartite signals mediate subcellular targeting of tail-anchored membrane proteins in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **278**, 8219–8223 (2003).
3. Kalbfleisch, T., Cambon, A. & Wattenberg, B. W. A bioinformatics approach to identifying tail-anchored proteins in the human genome. *Traffic* **8**, 1687–1694 (2007).
4. Kriechbaumer, V. *et al.* Subcellular distribution of tail-anchored proteins in *Arabidopsis*. *Traffic* **10**, 1753–1764 (2009).
5. Favaloro, V., Spasic, M., Schwappach, B. & Dobberstein, B. Distinct targeting pathways for the membrane insertion of tail-anchored (TA) proteins. *J. Cell Sci.* **121**, 1832–1840 (2008).
6. Favaloro, V., Vilardi, F., Schlecht, R., Mayer, M. P. & Dobberstein, B. Asna1/TRC40-mediated membrane insertion of tail-anchored proteins. *J. Cell Sci.* **123**, 1522–1530 (2010).
7. Schuldiner, M. *et al.* The GET complex mediates insertion of tail-anchored proteins into the ER membrane. *Cell* **134**, 634–645 (2008).
8. Stefanovic, S. & Hegde, R. S. Identification of a targeting factor for posttranslational membrane protein insertion into the ER. *Cell* **128**, 1147–1159 (2007).
9. Borgese, N. & Fasana, E. Targeting pathways of C-tail-anchored proteins. *Biochim. Biophys. Acta* **1808**, 937–946 (2011).
10. Sherrill, J., Mariappan, M., Dominik, P., Hegde, R. S. & Keenan, R. J. A conserved archaeal pathway for tail-anchored membrane protein insertion. *Traffic* doi:10.1111/j.1600-0854.2011.01229.x (3 July 2011).
11. Borgese, N. & Righi, M. Remote origins of tail-anchored proteins. *Traffic* **11**, 877–885 (2010).
12. Wang, F., Brown, E. C., Mak, G., Zhuang, J. & Denic, V. A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. *Mol. Cell* **40**, 159–171 (2010).
13. Mariappan, M. *et al.* A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature* **466**, 1120–1124 (2010).
14. Chang, Y.-W. *et al.* Crystal structure of Get4-Get5 complex and its interactions with Sgt2, Get3, and Ydj1. *J. Biol. Chem.* **285**, 9962–9970 (2010).
15. Jonikas, M. C. *et al.* Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* **323**, 1693–1697 (2009).
16. Mateja, A. *et al.* The structural basis of tail-anchored membrane protein recognition by Get3. *Nature* **461**, 361–366 (2009).
17. Bozkurt, G. *et al.* Structural insights into tail-anchored protein binding and membrane insertion by Get3. *Proc. Natl Acad. Sci. USA* **106**, 21131–21136 (2009).
18. Suloway, C. J., Chartron, J. W., Zaslaver, M. & Clemons, W. M. Jr. Model for eukaryotic tail-anchored protein binding based on the structure of Get3. *Proc. Natl Acad. Sci. USA* **106**, 14849–14854 (2009).
19. Yamagata, A. *et al.* Structural insight into the membrane insertion of tail-anchored proteins by Get3. *Genes Cells* **15**, 29–41 (2010).
20. Hu, J., Li, J., Qian, X., Denic, V. & Sha, B. The crystal structures of yeast Get3 suggest a mechanism for tail-anchored protein membrane insertion. *PLoS ONE* **4**, e8061 (2009).
21. Auld, K. L. *et al.* The conserved ATPase Get3/Arr4 modulates the activity of membrane-associated proteins in *Saccharomyces cerevisiae*. *Genetics* **174**, 215–227 (2006).
22. Leznicki, P., Clancy, A., Schwappach, B. & High, S. Bat3 promotes the membrane integration of tail-anchored proteins. *J. Cell Sci.* **123**, 2170–2178 (2010).
23. Schuldiner, M. *et al.* Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–519 (2005).
24. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425–431 (2010).
25. Chartron, J. W., Suloway, C. J. M., Zaslaver, M., a. & Clemons, W. M. Structural characterization of the Get4/Get5 complex and its interaction with Get3. *Proc. Natl Acad. Sci. USA* **107**, 12127–12132 (2010).
26. Renthal, R. Helix insertion into bilayers and the evolution of membrane proteins. *Cell. Mol. Life Sci.* **67**, 1077–1088 (2010).
27. Brambillasca, S. *et al.* Transmembrane topogenesis of a tail-anchored protein is modulated by membrane lipid composition. *EMBO J.* **24**, 2533–2542 (2005).
28. Sharma, A., Mariappan, M., Appathurai, S. & Hegde, R. S. *In vitro* dissection of protein translocation into the mammalian endoplasmic reticulum. *Methods Mol. Biol.* **619**, 339–363 (2010).
29. Brambillasca, S., Yabal, M., Makarow, M. & Borgese, N. Unassisted translocation of large polypeptide domains across phospholipid bilayers. *J. Cell Biol.* **175**, 767–777 (2006).
30. Panzner, S., Dreier, L., Hartmann, E., Kostka, S. & Rapoport, T. A. Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. *Cell* **81**, 561–570 (1995).
31. Rothblatt, J. A. & Meyer, D. I. Secretion in yeast: reconstitution of the translocation and glycosylation of alpha-factor and invertase in a homologous cell-free system. *Cell* **44**, 619–628 (1986).
32. Hansen, W., Garcia, P. D. & Walter, P. *In vitro* protein translocation across the yeast endoplasmic reticulum: ATP-dependent posttranslational translocation of the prepro-alpha-factor. *Cell* **45**, 397–406 (1986).

33. Gorlich, D. & Rapoport, T. A. Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell* **75,** 615–630 (1993).

34. Fons, R. D., Bogert, B. A. & Hegde, R. S. Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *J. Cell Biol.* **160,** 529–539 (2003).

35. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

36. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).

37. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).

Author Contributions A.M., M.D. and E.B. produced, purified and characterized recombinant Get1, Get2 (full length and fragments) and Get3. M.M. and R.S.H. performed the reconstitution experiments, including the substrate release and membrane insertion assays. A.M., M.D. and R.J.K. carried out crystallization and structure determination as well as the interaction analyses. R.S.H. and R.J.K. designed the project. M.M., R.S.H. and R.J.K. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Atomic coordinates and structure factors for *S. cerevisiae* Get3 in complex with Get1(21–104) and for Mg^{2+}-ADP•AlF_4^--bound *S. cerevisiae* Get3 in complex with Get2(1–38) have been deposited in the Protein Data Bank under accession codes 3ZS8 and 3ZS9, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.J.K. (bkeenan@uchicago.edu) or R.S.H. (hegde.science@gmail.com).

METHODS

Reagents and basic procedures. Antibodies against Get1 (residues 61–74) and Get2 (residues 2–12) were generated against synthetic peptides conjugated to KLH via terminal cysteines. Antibody against yeast Get3 was against the whole recombinant protein. Antibody production was by LAMPIRE Biological Laboratories. The antibodies against the 3F4 tag and Sec61β have been described previously[8]. The Sec61α antibody was a gift from Tom Rapoport (Harvard University). DeoxyBigCHAP (DBC) was obtained from Calbiochem. Yeast strains were from Open Biosystems collections and were provided by Tom Dever. The following lipids were obtained from Avanti Polar Lipids: egg phosphatidylcholine (PC), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (PE) and 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine-*N*-lissamine rhodamine B (rhodamine-PE). Each lipid was dissolved and stored in chloroform at −20 °C or −80 °C. Protease inhibitor cocktail was from Roche (EDTA-free Complete tablets) and dissolved as a ×25 stock in aqueous buffer just before use. *In vitro* translation, chemical crosslinking and immunoprecipitations were as described previously[8,13,28].

Preparation of proteins for functional analysis. The genes encoding full-length or cytosolic fragments of *S. cerevisiae* Get1, Get2 and Get3 were amplified by PCR from genomic DNA. Site-directed mutants were obtained by QuikChange mutagenesis (Stratagene). Unless otherwise noted, all constructs were subcloned into a pET28 derivative (Novagen) modified to incorporate a tobacco etch virus (TEV) protease cleavage site between an N-terminal 6×His tag and the polylinker. All constructs were verified by DNA sequencing.

Expression and purification of full-length Get3 (wild type and D57N) was carried out as described previously[16]. Full-length Get1 and Get2 (wild type and mutants) were expressed in *E. coli* Rosetta2/pLysS (Novagen) using the Overnight Express Autoinduction System 1 (Novagen). Cells were disrupted in buffer A (50 mM HEPES, pH 8.0, 500 mM NaCl, 10 mM imidazole, 5% glycerol) with 1 mM PMSF using a high-pressure microfluidizer (Avestin), and the insoluble pellet was isolated by centrifugation. This pellet was washed in buffer A, recentrifuged and solubilized for 1 h at 4 °C in buffer A containing 0.5% *n*-dodecyl-*N,N*-dimethylamine-*N*-oxide (LDAO). The detergent-soluble fraction was then subjected to nickel-affinity chromatography (Ni-NTA agarose, Qiagen) in buffer A containing 30 mM imidazole and 0.1% LDAO. Protein was eluted at ~1 mg ml^{-1} in buffer A containing 200 mM imidazole and 0.1% LDAO, and stored in aliquots at −80 °C. Protein concentrations were determined using calculated A_{280} extinction coefficients.

The cytosolic Get1 fragment (residues 21–104) was expressed for 3 h at 37 °C (wild type) or overnight at 25 °C (R73E mutant) in *E. coli* BL21(DE3)/pRIL (Novagen), following induction with 0.1 mM IPTG. Cells were disrupted in buffer B (50 mM Tris, pH 7.5, 500 mM NaCl, 10 mM imidazole, 5% glycerol, 5 mM β-mercaptoethanol) with 1 mM PMSF using a microfluidizer. After clearing by centrifugation, the supernatant was batch-purified by nickel-affinity chromatography. Protein was eluted in buffer B containing 200 mM imidazole, dialyzed into 10 mM Tris, pH 7.5, 250 mM NaCl and 40% glycerol, and then stored at −80 °C. This was typically followed by gel filtration (Superdex 200 10/300 GL, GE Healthcare) in 10 mM Tris, pH 7.5, and 200 mM NaCl. Fractions were pooled and stored in aliquots at −80 °C. Protein concentrations were determined using calculated A_{280} extinction coefficients.

The cytosolic Get2 fragment (residues 1–38 or 1–106; wild type and R17E) was expressed with an N- or C-terminal 6×His tag overnight at 25 °C and purified by nickel-affinity chromatography as described above for the Get1 fragment. After dialysis against 10 mM Tris, pH 7.5, and 200 mM NaCl, proteins were further purified by gel filtration in 10 mM Tris, pH 7.5, and 150 mM NaCl. Fractions were pooled, concentrated and stored in aliquots at −80 °C. Protein concentration was determined by BCA (Pierce).

Preparation of liposomes. The standard liposome mixture typically contained PC:PE:rhodamine-PE at a mass ratio of 8:1.9:0.1. Rhodamine-PE serves as a tracer to follow the lipid recovery. Lipid solutions were mixed in the above ratios as chloroform stocks, adjusted to 10 mM DTT and dried in a glass tube by centrifugation under vacuum (SpeedVac, Eppendorf) for 12 h. Lipid films were hydrated to a final concentration of 20 mg ml^{-1} in lipid buffer (50 mM HEPES-KOH, pH 7.4, 15% glycerol) and mixed end to end for 6 h at 25 °C with intermittent vortexing. The milky and uniform suspension was subjected to three freeze–thaw cycles (freeze in liquid nitrogen; thaw at 37 °C) and extruded at 65 °C 11 times through 100-nm polycarbonate membranes using an Avanti mini-extruder[27,29]. Single-use aliquots (100 µl) of the final clear liposome solution were flash-frozen in liquid nitrogen and stored at −80 °C.

Purification of recombinant targeting complex. The DNA template for the double-Strep-tagged human Sec61β was generated by PCR using a 5′ oligonucleotide that encodes the T7 promoter, start codon and tag. This template was transcribed and translated in RRL as described previously[28], but with 0.15 mg ml^{-1}

His-Get3 (added from a 20 mg ml^{-1} stock in 10 mM Tris-HCl, pH 7.5, 100 mM NaCl and 40% glycerol). A 2-ml translation reaction was diluted twofold with ice-cold column buffer (20 mM HEPES-KOH, pH 7.4, 100 mM potassium acetate, 2 mM magnesium acetate, 1 mM DTT) and centrifuged for 30 min at 540,960*g* in a TLA100.3 rotor at 4 °C. The post-ribosomal supernatant was bound to a 400-µl DEAE-Sepharose fast-flow column at 4 °C, washed with column buffer and eluted with a buffer containing 50 mM HEPES-KOH, pH 7.4, 320 mM potassium acetate, 7 mM magnesium acetate and 1 mM DTT. The elution was passed over 200 µl Strep-Tactin agarose (IBA, Germany) one to three times. After washing with four column volumes of Strep-Tactin buffer (50 mM HEPES-KOH, pH 7.4, 10% glycerol, 150 mM potassium acetate, 7 mM magnesium acetate, 1 mM DTT) at 4 °C, bound proteins were eluted with 5 × 50 µl Strep-Tactin buffer containing 10 mM Desthiobiotin (Novagen). The peak fractions, measured by counting radioactivity, were pooled. The final sample contained ~10,000 c.p.m. µl^{-1}. The concentration of Get3 in the final sample was estimated to be ~80 nM. Thus, the targeting complex in our typical preparation has a concentration of ~40 nM, assuming a 2:1 ratio of Get3 to Sec61β. This was either used immediately or frozen in aliquots in liquid nitrogen and stored at −80 °C. Targeting complexes containing the TMDs of rat VAMP2 and *S. cerevisiae* Sed5 in place of the Sec61β TMD were made similarly.

Insertion assay. Post-translational insertion assay was performed as described before[8], with the following minor modifications. For a standard reaction, 8 µl of purified targeting complex was mixed with 1 µl of ATP regenerating system (2 mM ATP, 10 mM creatine phosphate and 40 µg ml^{-1} creatine kinase) and 1 µl of yRMs, liposomes, reconstituted proteoliposomes or a matched buffer. ATP regenerating system was omitted in some reactions as indicated in the figure legends. After incubation at 32 °C for 30 min, the samples were treated with proteinase K (0.5 mg ml^{-1}) for 60 min on ice, and the protease digestion was terminated with 5 mM PMSF and transferred to 100 µl of boiling 1% SDS as described previously[8]. The protease-protected fragment was then immunoprecipitated using the 3F4 antibody directed against the C terminus of the Sec61β construct. Immunoprecipitated products were analysed by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and quantified by phosphor imaging.

Preparation of rough microsomes from yeast. Yeast microsomes were prepared by modifications of the methods previously described[30–32,38]. TAP-tagged Get1 (Open Biosystems) or Get deletion strains (gift from T. Dever) were grown at 30 °C to a density of $2A_{600}$ U in 1 l of YPD medium containing 2% glucose. Cells were collected by centrifugation at 3,000*g* for 5 min and washed twice with ice-cold distilled water. All subsequent steps were on ice or at 4 °C. The cell pellet was resuspended in 50 ml of homogenization buffer (20 mM HEPES-KOH, pH 7.4, 100 mM potassium acetate, 2 mM magnesium acetate) and centrifuged for 5 min at 3,000*g*. The resulting cell pellet was resuspended in homogenization buffer containing 2 mM DTT and protease inhibitor cocktail (Roche) at a concentration of 1 ml per gram of cell pellet. Pre-chilled glass beads were added (3 g ml^{-1} of suspension), and cell lysis was induced as follows: the tube was vigorously shaken up and down over a 50-cm path at ~1–2 cycles s^{-1} for three 1-min periods separated by 1 min chilling on ice. Approximately 50% of the cells were broken by this method as visualized by microscopy. The fluid phase was drained off through a fine nylon mesh into a JA17 tube and spun at 10,000*g* for 10 min. The post-mitochondrial supernatant was briefly centrifuged in an MLA80 rotor at 339,707*g* for 8 min. Each 2 ml of the clear supernatant was layered on 1 ml of 0.67 M sucrose cushion in homogenization buffer and centrifuged for 30 min in a TLA100.3 rotor at 265,070*g*. The resulting membrane pellet was resuspended in homogenization buffer containing 250 mM sucrose and 2 mM DTT to a final standard concentration of $100A_{280}$ (measured after solubilization in 1% SDS). At this concentration, 1 µl yRM is defined as two equivalents (equiv.). One litre of culture yielded about 2,400 equiv. Aliquots were frozen in liquid nitrogen and stored at −80 °C.

Depletion of Get1 and Get2 from microsomal extract. TAP-Get1 yRMs (1.5 ml, or 1,500 equiv.) were adjusted to 1% DBC in solubilization buffer (50 mM HEPES-KOH, pH 7.4, 500 mM potassium acetate, 5 mM magnesium acetate, 250 mM sucrose, 1 mM DTT and protease inhibitor cocktail). After 10 min incubation on ice, the detergent extract was centrifuged at 540,960*g* for 30 min in a TLA100.3 rotor at 4 °C. The supernatant (yRM extract) was incubated with 0.1 ml of IgG Sepharose (GE Healthcare) for 1 h at 25 °C. The unbound fraction was incubated with 0.1 ml of anti-Get2 antibodies coupled to protein-A agarose for 1 h at 25 °C. The flow-through was finally incubated with a mixture of 0.1 ml each of anti-Get1 and anti-Get2-antibodies coupled to protein-A agarose for 1 h at 25 °C. The flow-through from this column was used for reconstitution studies. It should be noted that a residual amount of the Get1/2 complex is sufficient to achieve the maximal insertion under *in vitro* conditions. Therefore, multiple rounds of depletion of the Get1/2 complex (with at least ~95% depletion) were necessary to fully deplete insertion activity. For purification of TAP-Get1 (and

associated proteins), the IgG Sepharose resin from above was washed with low-salt buffer (10 mM Tris, pH 7.4, 150 mM NaCl, 10% glycerol, 0.25% DBC and 1 mM DTT) and eluted with 70 U TEV-protease (Invitrogen) overnight at 4 °C. The TEV elution was adjusted to 2.5 mM CaCl$_2$, and incubated with calmodulin Sepharose (GE Healthcare) for 90 min at 4 °C. The beads were washed with low-salt buffer containing CaCl$_2$ and eluted with low-salt buffer containing 5 mM EGTA. The eluted proteins were precipitated with TCA and analysed by SDS–PAGE.

Reconstitution of proteoliposomes from microsome extracts. Following earlier methods[30,33,34], yRMs were adjusted to a concentration of 1 equiv. µl^{-1} in the following conditions: 50 mM HEPES-KOH, pH 7.4, 500 mM potassium acetate, 5 mM magnesium acetate, 250 mM sucrose, 1 mM DTT, 1% DBC and protease inhibitor cocktail. After 10 min on ice, the ribosomes were removed by centrifugation at 540,960g. for 30 min in a TLA100.3 rotor at 4 °C. Typically, 100 µl of this clarified yRM extract was mixed with 10 µl of liposomes (200 µg) and 50 mg of Biobeads SM2 (Bio-Rad). The Biobeads were prewashed extensively ahead of time with methanol and water. The mixture was incubated for 12–16 h with gentle overhead mixing at 4 °C. The fluid phase was separated from the beads, diluted with five volumes of ice-cold distilled water and sedimented in a TLA100.3 rotor in micro-test tubes at 304,290g for 30 min at 4 °C. The proteoliposomes were resuspended in 25 µl of membrane buffer (50 mM HEPES-KOH, pH 7.4, 100 mM potassium acetate, 5 mM magnesium acetate, 250 mM sucrose, and 1 mM DTT).

Reconstitution of proteoliposomes with purified proteins. The optimum method for reconstitution of purified Get1 or Get2 was empirically determined after testing various detergents and reconstitution methods (Supplementary Fig. 17) The precise method of reconstitution proved to be important for obtaining maximally functional proteoliposomes. The incorporation and activity of Get1 and Get2 varied with different detergents. Of those tested, DBC worked the best to achieve the maximal activity of Get1 and Get2. Every batch of DBC requires some degree of optimization with respect to the amount of Biobeads used for detergent removal. For a standard reconstitution reaction, 100 µl of reconstitution buffer (50 mM HEPES-KOH, pH 7.4, 500 mM potassium acetate, 5 mM magnesium acetate, 250 mM sucrose, 1 mM DTT, 0.25% DBC) was mixed with 10 µl of liposome (200 µg) and purified Get1 or Get2 at the desired concentration. For preparation of liposomes used as controls in the assays, purified proteins were omitted. This mixture was added to between 25 and 30 mg of Biobeads (optimized for each batch of DBC), and incubated with overhead mixing for 12 h at 4 °C. The fluid phase was separated and diluted with five volumes of ice-cold water. In some instances, the proteoliposomes were mixed with Get3 and incubated for 15 min at 25 °C, followed by 30 min at 4 °C with shaking, to allow binding. After dilution, the liposomes were sedimented in a TLA100.3 rotor in micro-test tubes at 304,290g for 30 min at 4 °C. The proteoliposomes were resuspended in 25 µl of membrane buffer as above. SDS–PAGE Coomassie staining and immunoblots were performed to assess the efficiency of protein incorporation; the rhodamine-PE served as a marker for lipid recovery. Typical recovery for Get1 and Get2 reconstitution was ~50%.

Multi-angle light scattering. The absolute molecular masses of individual proteins and complexes were measured by static multi-angle light scattering. Purified samples were injected onto a Superdex 200 HR 10/30 gel filtration column (GE Healthcare) equilibrated with 10 mM Tris, pH 7.5, 150 mM NaCl and 2 mM DTT. The purification system was coupled to an online, static, light-scattering detector (Dawn HELEOS II, Wyatt Technology), a refractive-index detector (Optilab rEX, Wyatt Technology) and a ultraviolet-light detector (UPC-900, GE Healthcare). Absolute weight-averaged molar masses were calculated using the ASTRA software (Wyatt Technology).

Receptor fragment binding assays. Gel-filtration-purified, 6×His-tagged Get1(21–104), Get2(1–106) and Get3 (wild type and D57N) proteins were labelled with amine-reactive succinimidyl esters of Alexa488 or Alexa594 (Invitrogen). Labelling reactions were carried out by incubating ~150 µM protein and ~600 µM dye for 1 h at room temperature (23 °C) in 100 mM NaHCO$_3$, pH 8.3, and 200 mM NaCl. After labelling, proteins were desalted and concentrated in Amicon Ultra filtration units (Millipore) to ~100 µM in 20 mM HEPES, pH 7.5, and 200 mM NaCl (receptor fragments) or 20 mM HEPES, pH 7.5, 200 mM NaCl and 2 mM DTT (Get3), and stored in aliquots at −80 °C. Protein concentration was determined using calculated A_{280} extinction coefficients after correcting for dye absorbance. Under these labelling conditions, we typically observed ~0.5–1.5 mol of dye per mole of protein.

Dissociation constants (K_d) were determined by titrating a fixed amount of labelled, nucleotide-free Get3 with labelled Get1c or Get2c. Fluorescence measurements were made in 96-well format using a Safire2 (Tecan) plate reader. Alexa594-labelled fragments were excited by fluorescence resonance energy transfer from Alexa488-labelled Get3 (wild type or D57N), using excitation and emission wavelengths of 495 and 615 nM, respectively. All experiments were

carried out in 150 µl of 50 mM HEPES, pH 7.5, 100 mM NaCl, 5 mM MgCl$_2$, 5% glycerol, 0.02% Tween20 and 2 mM DTT. Blank titrations were carried out in the absence of labelled Get3 and were subtracted from the respective titration curves obtained in the presence of labelled Get3. The difference curves were evaluated by nonlinear regression using the following quadratic binding equation: $\Delta Y = 0.5B_{max}/P(K_d + P + X - \sqrt{((K_d + P + X)^2 - 4PX)})$, where B_{max} is the amplitude, P is the total concentration of labelled Get3, and X is the total concentration of labelled Get1c or Get2c.

Chase titrations were carried out by measuring fluorescence resonance energy transfer between Alexa488-labelled Get3 (wild type or D57N) and Alexa594-labelled fragments in the presence of increasing concentrations of an unlabelled fragment or nucleotide. Blank titrations were performed in the absence of labelled Get3 and were subtracted from the respective titration curves obtained in the presence of labelled Get3. The difference curves were evaluated by nonlinear regression using the following equation: $\Delta Y = F_{end} + B_{max}P/(P + K_{d,labelled}(1 + X/K_d))$, where F_{end} is the signal at saturation, B_{max} is the amplitude of the signal change, P is the total concentration of labelled fragment, $K_{d,labelled}$ is the dissociation constant of the Get3 fragment complex, X is the total concentration of the unlabelled component and K_d is the dissociation constant of the unlabelled component.

Nucleotide binding assays. Fluorescence measurements were made in 96-well format using a Safire2 plate reader with excitation and emission wavelengths of 285 and 446 nm, respectively (Supplementary Fig. 18). All experiments were carried out with gel-filtration-purified, 6×His-tagged Get3 (D57N) in 150 µl of 50 mM HEPES, pH 7.5, 100 mM NaCl, 5 mM MgCl$_2$, 5% glycerol, 0.02% Tween20 and 2 mM DTT. The dissociation constant of mant-ATP was measured by incubating 1 µM of Get3 (D57N) with increasing concentrations of mant-ATP (Molecular Probes). Dissociation constants of unlabelled nucleotides were determined by incubating 1 µM Get3 (D57N) with 1 µM mant-ATP and chasing with increasing concentrations of the corresponding unlabelled nucleotide. In each case, blank titrations were performed in the absence of Get3 and were subtracted from titration curves obtained in the presence of labelled Get3. ATP and ADP concentrations were determined by absorbance ($\varepsilon^{259} = 15,400 \text{ M}^{-1} \text{ cm}^{-1}$). Dissociation constants were determined by curve fitting as described above.

Tail-anchored substrate release assay. Get3–substrate complexes were assembled by *in vitro* translation in a phenyl- and DEAE-Sepharose-depleted RRL[10] supplemented with 6×His–Get3 at ~2 µg ml^{-1}. This translation extract lacks endogenous TA binding proteins (particularly TRC40 and Bag6). Translation of the TA substrate in this system was verified to result in Get3–substrate complexes by crosslinking, and was functional as judged by Get1/2-dependent insertion (data not shown). Complexes generated by this method were mixed with the fragments or proteoliposomes as indicated in the figure legends, incubated for 30 min at 32 °C and subjected to crosslinking with disuccinimidyl suberate as described previously[8]. The samples were denatured in 1% SDS, diluted tenfold in 1% Triton X-100 buffer and subjected to pull-downs of 6×His–Get3 with immobilized Co^{2+} bound to chelating Sepharose (GE). The Get3–substrate crosslink was visualized by autoradiography.

Preparation of Get3 receptor fragment complexes for crystallization. The gene encoding native, full-length *S. cerevisiae* Get3 was subcloned into pET19b (Novagen). For co-expression with N-terminal 6×His-tagged Get1(21–104), plasmids were co-transformed into *E. coli* BL21(DE3)/pRIL (Novagen). Proteins were expressed at 37 °C for 3 h by induction with 0.1 mM IPTG after the cells reached an A_{600} of ~0.6. Cells were disrupted and purified by nickel-affinity chromatography as described above for the Get1 and Get2 fragments. Protein was eluted in buffer B containing 200 mM imidazole, and then dialysed into 10 mM Tris, pH 7.5, 100 mM NaCl, 2 mM DTT and 40% glycerol. This was followed by cleavage with 6×His-tagged TEV protease and removal of residual uncleaved Get1 fragments and the 6×His-tagged TEV protease by subtractive Ni-NTA purification. Finally, the complex was separated from excess Get1 fragments by gel filtration. Fractions were pooled, concentrated to ~10 mg ml^{-1} in 10 mM Tris, pH 7.5, 100 mM NaCl and 2 mM DTT, and stored in aliquots at −80 °C.

Co-expression of native Get3 and N-terminal 6×His-tagged Get2(1–106) or Get2(1–38) was performed as above, except that proteins were expressed at 25 °C for 6–8 h after induction. Following cell lysis and purification by nickel-affinity chromatography, the protein was dialysed into 10 mM Tris, pH 7.5, 200 mM NaCl and 2 mM DTT. This was followed by cleavage with 6×His-tagged TEV protease and subtractive Ni-NTA purification. Finally, the complex was separated from excess Get2 fragments by gel filtration. Fractions were pooled, concentrated to ~15–20 mg ml^{-1} in 10 mM Tris, pH 7.5, 150 mM NaCl and 2 mM DTT, and stored in aliquots at −80 °C.

Crystallization. Crystals of *S. cerevisiae* Get1(21–104) in complex with *S. cerevisiae* Get3 were grown at room temperature using hanging-drop vapour diffusion by mixing equal volumes of a protein solution with a reservoir solution containing 0.2 M K/Na tartrate, 16% PEG 3350, 0.1 M HEPES, pH 7.2, and 6% polypropylene

glycol P400. Crystals were cryoprotected in mother liquor supplemented with 20% ethylene glycol, and flash-frozen in liquid nitrogen.

Crystals of *S. cerevisiae* Get2(1–38) in complex with *S. cerevisiae* Get3 and ADP•AlF$_4^-$ were grown at room temperature using hanging-drop vapour diffusion by mixing equal volumes of a protein solution containing 2 mM ADP, 2 mM MgCl$_2$, 2 mM AlCl$_3$ and 8 mM NaF with a reservoir solution containing 30% PEG 3350, 0.3 M ammonium acetate and 0.1 M Bis-Tris, pH 6.0. Crystals were briefly soaked in mother liquor supplemented with 20% ethylene glycol and flash-frozen in liquid nitrogen.

Structure determination and refinement. All data were collected at 100 K at APS beamline 21-IDG ($\lambda = 0.97856$ Å) and processed using HKL2000 (HKL Research). Data collection and refinement statistics are listed in Supplementary Table 1.

The structure of the Get1(21–104) complex with Get3 was determined to a resolution of 3.0 Å by molecular replacement with PHASER[35], using the open-dimer (nucleotide-free) form of *S. cerevisiae* Get3 (PDB ID, 3H84[20]; with the α-helical subdomain removed) as the search model. No solution could be obtained using the closed-dimer form of *S. cerevisiae* Get3 as the search model. Clear density was observed for the helical Get1 fragment and portions of the Get3 α-helical subdomain in the initial electron density maps. Model building and refinement were carried out in PHENIX[36] and COOT[37]. The final model contains one Get3 homodimer (chains A and B), two Get1 fragments (chains C and D) and one zinc atom, and was refined to an *R*-factor of 22.4% ($R_{\text{free}} = 28.2\%$). Most (94.3%) of the residues are in favoured regions of the Ramachandran plot, and 0.9% are outliers. Side-chain density is generally weakest in the α-helical subdomains, and no interpretable density was observed for residues 1–4, 97–134, 155–157, 198–219, 280–284 and 352–354 in chain A; 1–4, 99–125, 191–210,

280–284 and 352–354 in chain B; 21–35 and 103–104 in chain C; and 21–36 and 99–104 in chain D.

The structure of the Get2(1–38) complex with Get3 was determined to a resolution of 2.1 Å by molecular replacement with PHASER using a monomer of *S. cerevisiae* Get3 (PDB ID, 2WOJ[16]; with the α-helical subdomain and ligands removed) as the search model. Density for the two helices of Get2(1–38) and portions of the Get3 α-helical subdomain was clearly visible in the initial electron density maps. Model building and refinement were carried out in PHENIX and COOT. The final model contains one Get3 homodimer (chains A and B), two Get2 fragments (chains C and D), two Mg^{2+}-ADP•AlF$_4^-$ complexes, one zinc atom and 231 water molecules, and was refined to an *R*-factor of 18.8% ($R_{\text{free}} = 23.3\%$). Again, most (98.0%) of the residues are in favoured regions of the Ramachandran plot, and 0.8% are outliers. No interpretable electron density was observed for residues 1–4, 101–126, 188–211, 280–284 and 353–354 in chain A; 1–3, 102–125, 154–158, 199–211, 280–282 and 351–354 in chain B; 1–3 and 35–38 in chain C; and 1–3 in chain D.

Miscellaneous. SDS–PAGE was done with 15% Tris-glycine or 12% Tris-tricine gels. Quantification was by phosphor imaging using a Typhoon system with accompanying software. Most images for the figures were generated by exposure to Kodak MR X-ray film. Films were digitized by scanning. Structure figures were generated with Pymol[39] and all figures were assembled using Adobe Photoshop and Illustrator.

38. Deshaies, R. J. & Schekman, R. SEC62 encodes a putative membrane protein required for protein translocation into the yeast endoplasmic reticulum. *J. Cell Biol.* **109,** 2653–2664 (1989).
39. DeLano, W. L. *PyMOL Molecular Viewer* 〈http://www.pymol.org〉 (2002).

LETTER

An extremely primitive star in the Galactic halo

Elisabetta Caffau[1,2], Piercarlo Bonifacio[2], Patrick François[2,3], Luca Sbordone[1,2,4], Lorenzo Monaco[5], Monique Spite[2], François Spite[2], Hans-G. Ludwig[1,2], Roger Cayrel[2], Simone Zaggia[6], François Hammer[2], Sofia Randich[7], Paolo Molaro[8] & Vanessa Hill[9]

The early Universe had a chemical composition consisting of hydrogen, helium and traces of lithium[1]; almost all other elements were subsequently created in stars and supernovae. The mass fraction of elements more massive than helium, Z, is known as 'metallicity'. A number of very metal-poor stars has been found[2,3], some of which have a low iron abundance but are rich in carbon, nitrogen and oxygen[4-6]. For theoretical reasons[7,8] and because of an observed absence of stars with $Z < 1.5 \times 10^{-5}$, it has been suggested that low-mass stars cannot form from the primitive interstellar medium until it has been enriched above a critical value of Z, estimated to lie in the range 1.5×10^{-8} to 1.5×10^{-6} (ref. 8), although competing theories claiming the contrary do exist[9]. (We use 'low-mass' here to mean a stellar mass of less than 0.8 solar masses, the stars that survive to the present day.) Here we report the chemical composition of a star in the Galactic halo with a very low Z ($\leq 6.9 \times 10^{-7}$, which is 4.5×10^{-5} times that of the Sun[10]) and a chemical pattern typical of classical extremely metal-poor stars[2,3]—that is, without enrichment of carbon, nitrogen and oxygen. This shows that low-mass stars can be formed at very low metallicity, that is, below the critical value of Z. Lithium is not detected, suggesting a low-metallicity extension of the previously observed trend in lithium depletion[11]. Such lithium depletion implies that the stellar material must have experienced temperatures above two million kelvin in its history, given that this is necessary to destroy lithium.

The Galactic halo star SDSS J102915+172927, the object of this Letter, has been observed with the X-Shooter[12] and UVES[13] spectrographs at the Very Large Telescope (VLT), operated by the European Southern Observatory in Chile. Its properties are as follows: right ascension, 10 h 29 min 15.15 s, declination, $+17° 29' 28''$ at equinox 2000; g band magnitude 16.92, $(g - z) = 0.59$ mag, and, after correction for interstellar reddening, $(g-z)_0 = 0.53$ mag. A portion of the spectra in the region of the Ca II K line is shown in Fig. 1. We have computed and used theoretical model atmospheres and spectrum synthesis techniques to derive the chemical abundances provided in Table 1. The chemical signatures are consistent with metal production by ordinary core-collapse supernovae[14]. The derived abundances, coupled with the upper limits on carbon and nitrogen, imply $Z \leq 6.9 \times 10^{-7}$. This number takes into account the typical 'excess' of the α-element oxygen, $[O/Fe] = +0.6$. (Here $[A/B] = \log(N_A/N_B) - \log(N_A/N_B)_\odot$ for the number N of atoms of elements A and B, and subscript $\odot$ indicates the solar value.) Our analysis has been performed assuming local thermodynamic equilibrium (LTE); however, further work is necessary to assess the role of departures from LTE, especially for molecules. The estimate of non-LTE effects on magnesium[15] is about $+0.4$ dex, which translates in a change of $+0.2 \times 10^{-7}$ in Z.

It has been suggested that the primary discriminants between the formation of only massive stars (as in stellar population III) and of both massive and low-mass stars (as in stellar populations II and I) are the abundances of carbon and oxygen[7]. In this scenario, these elements can provide efficient cooling of the protostellar clouds in the primitive interstellar medium through the fine structure lines of ionized carbon and neutral oxygen. A suitable combination of the carbon and oxygen abundances is called the transition discriminant[16] ($D = \log_{10}(10^{[C/H]} + 0.3 \times 10^{[O/H]})$), and low-mass star formation is believed to occur only if $D \geq -3.5$. From the abundances in Table 1 and the assumption $[O/Fe] = +0.6$ we have $D \leq -4.2$ for SDSS J102915+172927, which places it in the 'forbidden zone' of the theory. If, instead of taking the upper limit on the carbon abundance, we assume that the carbon abundance (derived from the three-dimensional (3D) analysis) scales with the iron abundance, as found in other metal-poor stars[3], we have $D \leq -4.4$. Our measurement cannot rule out the above-mentioned theoretical scenario[7,16], but it strongly supports the idea that, at least in some cases, low-mass stars can also form at lower carbon and oxygen abundances than the current estimates for the critical values.

The complete absence of the neutral lithium (Li I) resonance doublet at 670.7 nm, both in UVES and X-Shooter spectra, is remarkable. In
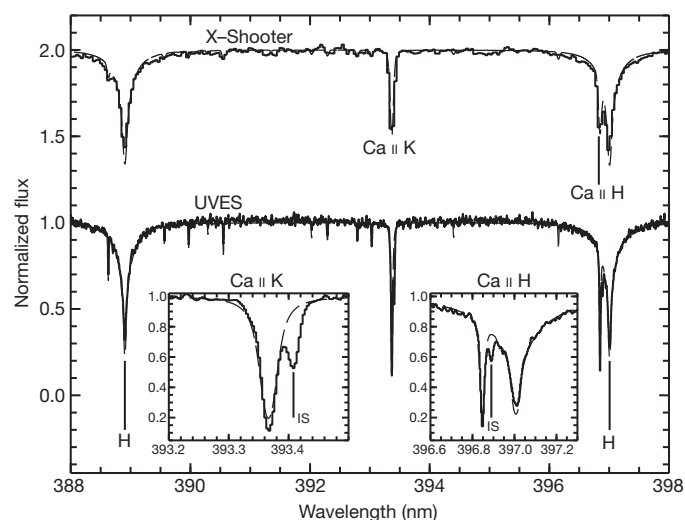
Figure 1 | Observed spectra of SDSS J102915+172927. The spectral region of the Ca II H and K lines is shown (solid lines), compared to synthetic spectra (long dashed lines) computed with a global metallicity of $Z = 1.1 \times 10^{-6}$ and solar proportions of all elements, except for α-elements that are enhanced by 0.4 dex over iron. Main figure: top trace, the X-Shooter spectrum (shifted vertically by one unit for clarity); bottom trace, the UVES spectrum. The spectra have been normalized to 1 in the continuum. Insets, magnified views of the Ca II K line (left) and the Ca II H line (right). The absorption due to interstellar gas is clearly detectable both in K and H Ca II lines (labelled as 'IS' in the figure), and two hydrogen lines, H8 and Hε (labelled as 'H' in the figure), are visible. The measured radial velocity is -34.5 ± 1.0 km s^{-1}. We computed a Galactic orbit from the kinematic data and a distance of 1.27 ± 0.15 kpc, estimated from the photometry, confirming that the star belongs to the Galactic halo.

[1]Zentrum für Astronomie der Universität Heidelberg, Landessternwarte, Königstuhl 12, 69117 Heidelberg, Germany. [2]GEPI, Observatoire de Paris, CNRS, Université Paris Diderot, Place Jules Janssen, 92190 Meudon, France. [3]UPJV, Université de Picardie Jules Verne, 33 Rue St Leu, F-80080 Amiens, France. [4]Max-Planck Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, 85741 Garching, Germany. [5]European Southern Observatory, Casilla 19001, Santiago, Chile. [6]Istituto Nazionale di Astrofisica, Osservatorio Astronomico di Padova Vicolo dell'Osservatorio 5, 35122 Padova, Italy. [7]Istituto Nazionale di Astrofisica, Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125 Firenze, Italy. [8]Istituto Nazionale di Astrofisica, Osservatorio Astronomico di Trieste, Via Tiepolo 11, 34143 Trieste, Italy. [9]Université de Nice-Sophia Antipolis, Observatoire de la Côte d'Azur, CNRS, Laboratoire Cassiopée, Boulevard de l'Observatoire, 06300 Nice, France.

Table 1 | Abundances in SDSS J102915+172927

Element	A(X), 3D	[X/H], 3D	[X/Fe], 3D	[X/H], 1D	Number of lines	A(X)$_\odot$
C	≤4.2	≤−4.3	≤+0.7	≤−3.8	G band	8.50
N	≤3.1	≤−4.8	≤+0.2	≤−4.1	NH band	7.86
Mg I	2.95	−4.59 ± 0.10	+0.40	−4.68 ± 0.08	4	7.54
Si I	3.25	−4.27 ± 0.10	+0.72	−4.27 ± 0.10	1	7.52
Ca I	1.53	−4.80 ± 0.10	+0.19	−4.72 ± 0.10	1	6.33
Ca II	1.48	−4.85 ± 0.11	+0.14	−4.71 ± 0.11	3	6.33
Ti II	0.14	−4.76 ± 0.11	+0.23	−4.75 ± 0.11	6	4.90
Fe I	2.53	−4.99 ± 0.12	+0.00	−4.73 ± 0.13	44	7.52
Ni I	1.35	−4.88 ± 0.11	+0.11	−4.55 ± 0.14	10	6.23
Sr II	≤−2.28	≤−5.2	≤−0.21	≤−5.1	1	2.92

These elemental abundances are derived from our UVES spectra. The last column provides the adopted solar abundances on the scale $A(X) = \log_{10}(X/H) + 12$. The atmospheric parameters adopted are effective temperature $T_{eff} = 5{,}811$ K, $\log g = 4.0$ (where g is acceleration due to gravity, in cm s^{-2}, at the surface), and microturbulent velocity 1.5 km s^{-1}. T_{eff} was derived from the $(g - z)_0$ colour–T_{eff} calibration[26]. The combination of photometric and reddening uncertainties causes an uncertainty on T_{eff} of 100 K, the corresponding uncertainty on [Fe/H] is 0.06 dex. We cross-checked the values of T_{eff} with a fit of the Hα wings which provided the same effective temperature within 10 K. The surface gravity has been fixed from the Balmer jump, as measured by the $(u - g)$ colour. Other gravity indicators, such as the calcium ionization equilibrium and the wings of higher-order Balmer lines, are consistent with this choice. The uncertainty on the surface gravity is about 0.2 dex, and we can robustly exclude a surface gravity $\log g = 3.0$ or lower, thus excluding that the star could be on the horizontal branch. We computed synthetic spectra with the SYNTHE code[27] and a one-dimensional (1D) model atmosphere was computed with the ATLAS 9 code[27]. These synthetic spectra were used to perform line-profile fitting for all the measurable features. The 3D corrections were computed using a 3D model atmosphere from the CIFIST grid[28] with $T_{eff} = 5{,}850$ K, $\log g = 4.0$ and metallicity 2.7×10^{-5}. We were able to measure the abundances of only some α-elements (Mg, Ca, Si, Ti) and two iron peak elements (Fe and Ni). The derived iron abundance is [Fe/H] = −4.99 (see Table 1; the 3D-corrected abundances in columns 2, 3 and 4 should be used, the 1D abundances in column 5 are given for reference only). The α-elements are slightly enhanced relative to iron, [Mg/Fe] = +0.4. The Sr II line at 407.8 nm is not convincingly detected, giving an upper limit [Sr/Fe] ≤ −0.21, which is compatible with the general pattern of low [Sr/Fe] found in extremely metal-poor stars[29]. The upper limits on carbon and nitrogen are derived by fitting the molecular bands of CH (G band) and NH, at 430 nm and 336 nm, respectively. Unfortunately no measurement of oxygen is possible in the available spectral range, neither from atomic nor from molecular lines, but there is no reason to suspect that a star not enhanced in Mg, C and N should be over-abundant in oxygen.

fact most of the 'warm' (effective temperature $T_{eff} > 5{,}700$ K) metal-poor dwarf stars display a constant abundance of Li, the so-called Spite plateau[11,17]. From the signal-to-noise ratio in the UVES spectrum of SDSS J102915+172927, we derive an upper limit for the Li abundance, $A(Li) < 1.1$ (at 5σ). In Fig. 2 we show the Spite plateau as a function of the carbon abundance, as well as a function of the iron abundance, which we use in turn as a proxy for Z. The sample of stars is composed of those with a normal carbon abundance[3,18–20] and the carbon-rich, iron-poor subgiant HE 1327−2326 (ref. 4). The pictures emerging from Fig. 2a and b show the same morphology, with the exception of star HE 1327−2326, which has [Fe/H] lower than all the others, but has [C/H] comparable to many other stars in the sample. It is noteworthy that the only two unevolved stars with [Fe/H] < −4.5 have no detectable Li.

The most straightforward interpretation of the Spite plateau is that the Li observed in the plateau stars is the Li produced in the Big Bang[17]. The theoretical primordial Li abundance[1] is a factor of 2–3 larger than the value observed on the Spite plateau. A number of explanations of this discrepancy have been proposed, which range from stellar phenomena, such as atomic diffusion[21], to new physics leading to a different Big Bang nucleosynthesis[22]. Our upper limit implies that the Li abundance of SDSS J102915+172927 is far below the value of the Spite plateau. At extremely low metallicities, the Spite plateau displays a 'meltdown'[11], that is, an increased scatter and a lower mean Li abundance. This meltdown is clearly seen in the two components of the extremely metal-poor binary system CS 22876-32, which show a different Li content[19]. The primary is on the Spite plateau, whereas the secondary is below it, at $A(Li) = 1.8$. The reasons for this meltdown are not understood. It has been suggested[11] that a Li depletion mechanism, whose efficiency depends on metallicity and temperature, could explain the observations. If this were the case, the Li abundance in SDSS J102915+172927 would result from efficient Li depletion due to a combination of extremely low metallicity and relatively low temperature. For completeness, we mention that there are a small number of known stars which have a metallicity and effective temperature similar to that of other stars on the Spite plateau, but where the Li doublet is
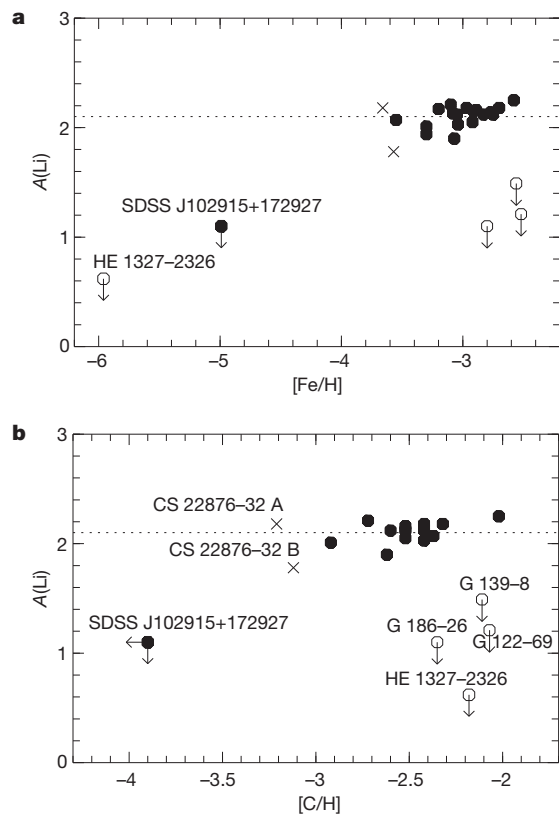
Figure 2 | Lithium abundance of SDSS J102915+172927 compared to that of other metal-poor stars. a, b, The Spite plateau is shown as a function of iron abundance, [Fe/H] (**a**), and of carbon abundance, [C/H] (**b**). We use carbon and iron in turn as proxies of the metallicity Z. The upper limits for SDSS J102915+172927 are from the present work. The other filled black circles are from refs 3 and 18, the upper limit for HE 1327−2326 is from ref. 4. The other open black circles are the upper limits and are mentioned below. The Li measurements for the binary system CS 22876–32 are from ref. 19, the upper limits for the three well known Li-depleted dwarfs, G 122-69, G 139-8 and G 186-26, are from ref. 20. For these three last stars, as well as for CS 22876–32, we have no measurement of the C abundances, and we therefore assumed that C scales with iron as in the rest of the sample[3]. The precise placement of these stars along the abscissa in this diagram is of no consequence for the present discussion.

not detected. The fact that such stars are found for different values of [Fe/H] and [C/H] suggests that Li-depletion is independent of either. It has been suggested that Li-depleted stars could have a common origin with blue stragglers[23], an interpretation that has been reinforced by the discovery that these stars are also depleted in beryllium[24].

Stars similar to SDSS J102915+172927 are probably not very rare. Only 30% of the whole SDSS survey area was accessible to our VLT observations. We identified 2,899 potentially extreme stars with metallicity less than $Z \leq 1.1 \times 10^{-5}$ in Data Release 7[25]. Among those observable with the VLT, we performed a subjective selection of the most promising candidates; of these, we observed six in our X-Shooter programme, resulting in one detection. Depending on the subjective bias we attribute to the last selection step, we expect 5–50 stars of similar or even lower metallicity than SDSS J102915+172927 to be found among the candidates accessible from the VLT, and many more in the whole SDSS sample.

1. Iocco, F., Mangano, G., Miele, G., Pisanti, O. & Serpico, P. D. Primordial nucleosynthesis: from precision cosmology to fundamental physics. *Phys. Rep.* **472,** 1–76 (2009).
2. Cayrel, R. *et al.* First stars V — Abundance patterns from C to Zn and supernova yields in the early Galaxy. *Astron. Astrophys.* **416,** 1117–1138 (2004).

3. Bonifacio, P. *et al.* First stars XII. Abundances in extremely metal-poor turnoff stars, and comparison with the giants. *Astron. Astrophys.* **501,** 519–530 (2009).

4. Frebel, A., Collet, R., Eriksson, K., Christlieb, N. & Aoki, W. HE 1327–2326, an unevolved star with [Fe/H]<−5.0. II. New 3D–1D corrected abundances from a Very Large Telescope UVES spectrum. *Astrophys. J.* **684,** 588–602 (2008).

5. Christlieb, N. *et al.* A stellar relic from the early Milky Way. *Nature* **419,** 904–906 (2002).

6. Norris, J. E. *et al.* HE 0557–4840: ultra-metal-poor and carbon-rich. *Astrophys. J.* **670,** 774–788 (2007).

7. Bromm, V. & Loeb, A. The formation of the first low-mass stars from gas with low carbon and oxygen abundances. *Nature* **425,** 812–814 (2003).

8. Schneider, R., Ferrara, A., Salvaterra, R., Omukai, K. & Bromm, V. Low-mass relics of early star formation. *Nature* **422,** 869–871 (2003).

9. Nakamura, F. & Umemura, M. On the initial mass function of Population III stars. *Astrophys. J.* **548,** 19–32 (2001).

10. Caffau, E., Ludwig, H.-G., Steffen, M., Freytag, B. & Bonifacio, P. Solar chemical abundances determined with a CO5BOLD 3D model atmosphere. *Sol. Phys.* **268,** 255–269 (2011).

11. Sbordone, L. *et al.* The metal-poor end of the Spite plateau. I. Stellar parameters, metallicities, and lithium abundances. *Astron. Astrophys.* **522,** A26 (2010).

12. D'Odorico, S. *et al.* X-shooter UV- to K-band intermediate-resolution high-efficiency spectrograph for the VLT: status report at the final design review. *Proc. SPIE* **6269,** 626933 (2006).

13. Dekker, H., D'Odorico, S., Kaufer, A., Delabre, B. & Kotzlowski, H. Design, construction, and performance of UVES, the echelle spectrograph for the UT2 Kueyen Telescope at the ESO Paranal. *Proc. SPIE* **4008,** 534–545 (2000).

14. Chieffi, A. & Limongi, M. The explosive yields produced by the first generation of core collapse supernovae and the chemical composition of extremely metal poor stars. *Astrophys. J.* **577,** 281–294 (2002).

15. Andrievsky, S. M. *et al.* Non-LTE abundances of Mg and K in extremely metal-poor stars and the evolution of [O/Mg], [Na/Mg], [Al/Mg], and [K/Mg] in the Milky Way. *Astron. Astrophys.* **509,** A88 (2010).

16. Frebel, A., Johnson, J. L. & Bromm, V. Probing the formation of the first low-mass stars with stellar archaeology. *Mon. Not. R. Astron. Soc.* **380,** L40–L44 (2007).

17. Spite, M. & Spite, F. Lithium abundance at the formation of the Galaxy. *Nature* **297,** 483–485 (1982).

18. Bonifacio, P. *et al.* First stars VII — Lithium in extremely metal poor dwarfs. *Astron. Astrophys.* **462,** 851–864 (2007).

19. González Hernández, J. I. *et al.* First stars XI. Chemical composition of the extremely metal-poor dwarfs in the binary CS 22876–032. *Astron. Astrophys.* **480,** 233–246 (2008).

20. Norris, J. E., Ryan, S. G., Beers, T. C. & Deliyannis, C. P. Extremely metal-poor stars. III. The Li-depleted main-sequence turnoff dwarfs. *Astrophys. J.* **485,** 370–379 (1997).

21. Richard, O., Michaud, G. & Richer, J. Implications of WMAP observations on Li abundance and stellar evolution models. *Astrophys. J.* **619,** 538–548 (2005).

22. Jedamzik, K. & Pospelov, M. Big Bang nucleosynthesis and particle dark matter. *N. J. Phys.* **11,** 105028 (2009).

23. Ryan, S. G., Gregory, S. G., Kolb, U., Beers, T. C. & Kajino, T. Rapid rotation of ultra-Li-depleted halo stars and their association with blue stragglers. *Astrophys. J.* **571,** 501–511 (2002).

24. Boesgaard, A. M. Beryllium in ultra-lithium-deficient halo stars: the blue straggler connection. *Astrophys. J.* **667,** 1196–1205 (2007).

25. Abazajian, K. N. *et al.* The seventh data release of the Sloan Digital Sky Survey. *Astrophys. J.* **182** (Suppl.), 543–558 (2009).

26. Ludwig, H.-G. *et al.* Extremely metal-poor stars from the SDSS. *Phys. Scripta T* **133,** 014037 (2008).

27. Kurucz, R. L. ATLAS12, SYNTHE, ATLAS9, WIDTH9, et cetera. *Mem. Soc. Astron. Ital.,* Suppl. **8,** 14–24 (2005).

28. Ludwig, H.-G. *et al.* The CIFIST 3D model atmosphere grid. *Mem. Soc. Astron. Ital.* **80,** 711–714 (2009).

29. François, P. *et al.* First stars. VIII. Enrichment of the neutron-capture elements in the early Galaxy. *Astron. Astrophys.* **476,** 935–950 (2007).

LETTER

Chronological evidence that the Moon is either young or did not have a global magma ocean

Lars E. Borg[1], James N. Connelly[2], Maud Boyet[3] & Richard W. Carlson[4]

Chemical evolution of planetary bodies, ranging from asteroids to the large rocky planets, is thought to begin with differentiation through solidification of magma oceans many hundreds of kilometres in depth[1–3]. The Earth's Moon is the archetypical example of this type of differentiation. Evidence for a lunar magma ocean is derived largely from the widespread distribution, compositional and mineralogical characteristics, and ancient ages inferred for the ferroan anorthosite (FAN) suite of lunar crustal rocks. The FANs are considered to be primary lunar flotation-cumulate crust that crystallized in the latter stages of magma ocean solidification. According to this theory, FANs represent the oldest lunar crustal rock type[2–4]. Attempts to date this rock suite have yielded ambiguous results, however, because individual isochron measurements are typically incompatible with the geochemical make-up of the samples, and have not been confirmed by additional isotopic systems[5–9]. By making improvements to the standard isotopic techniques, we report here the age of crystallization of FAN 60025 using the ^{207}Pb–^{206}Pb, ^{147}Sm–^{143}Nd and ^{146}Sm–^{142}Nd isotopic systems to be 4,360 ± 3 million years. This extraordinarily young age requires that either the Moon solidified significantly later than most previous estimates or the long-held assumption that FANs are flotation cumulates of a primordial magma ocean is incorrect. If the latter is correct, then much of the lunar crust may have been produced by non-magma-ocean processes, such as serial magmatism[10].

To constrain the age of a FAN using multiple chronometers we (1) obtained a unique mafic-mineral-rich sample of FAN 60025 found by examining the lunar collection at the Johnson Space Center, (2) developed a new methodology for Pb–Pb chronometry in which samples are washed numerous times to remove Pb contamination, and (3) produced a 99.988% pure ^{150}Nd spike that allowed parent/daughter ratios and isotopic compositions for both ^{146}Sm–^{142}Nd and ^{147}Sm–^{143}Nd chronometry to be measured in the same mineral fractions. This FAN was selected because it is coarse-grained and contains very low abundances of siderophile elements, indicating that it has been minimally altered by impact processes[11,12]. Furthermore, it has experienced a minimal flux of thermal neutrons, eliminating the need for corrections on Sm and Nd isotopic measurements[5,13]. Isotopic measurements were completed using large (about 500 mg) fractions of mafic minerals (mostly pyroxene) and plagioclase (see Supplementary Information).

The Pb–Pb and ^{147}Sm–^{143}Nd ages for 60025 are concordant and have a weighted average value of 4,360 ± 3 million years (Myr). This is the first study in which a single FAN has yielded consistent ages from multiple isotopic systems, and strongly suggests that the ages record the time at which the rock crystallized. The Pb–Pb age was determined on sequential dissolutions of a 105.4-mg pyroxene fraction. Four of five dissolution steps of this pyroxene mineral split define a line corresponding to an age of 4,359.2 ± 2.4 Myr (Fig. 1). Several mineral and whole-rock fractions analysed previously[6,14,15] also plot on the 4,359.2 ± 2.4 Myr line but show limited Pb isotopic variation. We interpret the line defined by four

dissolution steps of the pyroxene mineral splits to represent a mixture of initial Pb and radiogenic Pb that evolved in a closed system. We note that the Px-2 (L3) fraction has a ^{204}Pb/^{206}Pb ratio of 0.000774, indicating that it is dominated by radiogenic Pb, and consequently defines the y intercept on Fig. 1. This is by far the most radiogenic Pb measured in 60025 so far. For example, in a previous detailed study of 60025, the most radiogenic Pb reported was ^{204}Pb/^{206}Pb = 0.0151 (ref. 6). The radiogenic ^{207}Pb/^{206}Pb ratio defined by the y intercept is thus well constrained and corresponds to the age of last closure of this system.

The washing and dissolution procedures used to define the Pb–Pb isochron were developed using small (about 10 mg) fractions of pyroxene and plagioclase as test samples. Although the Pb isotopic compositions of these fractions were analysed to assess the efficiency of the procedure, relatively large blank corrections precluded their use in defining the age of 60025. The isotopic compositions of the washes demonstrate that most fall below the isochron towards modern terrestrial Pb. The Px-2 (L2) 1 M hydrofluoric acid (HF) acid fraction also lies towards terrestrial Pb on Fig. 1. Sequential washing and dissolution of primitive chondritic meteorites demonstrates that the first HF digestion step often liberates a large amount of terrestrial Pb contamination[16]. This probably reflects dissolution of small particles, such as dust, in the weak HF acid, which were not removed in the washing steps. It is therefore not surprising that the Px-2 (L2) fraction lies off the isochron towards terrestrial contamination, which justifies removing this fraction from the age calculation. However, including the Px-2 (L2) fraction in the regression only changes the calculated age by about 10 Myr. A few fractions fall above the isochron as well, and are interpreted to have a well known, if poorly understood, lunar Pb contaminant[14,15,17] (see Supplementary Information).
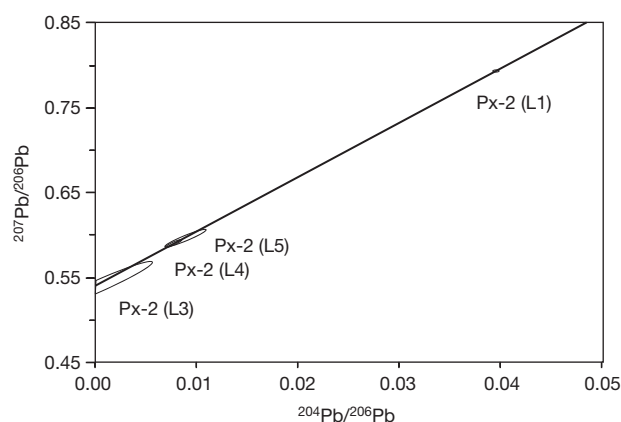
Figure 1 | Pb–Pb isochron diagram. For FAN 60025, an age of 4,359.2 ± 2.4 Myr is defined by sequential dissolutions of a 105.4-mg split of the pyroxene mineral fraction. Uncertainties are ±2σ mean of population of mass spectrometry ratios plus 50% uncertainty associated with a blank. The mean squared weighted deviation (MSWD) is 1.6.

[1]Chemical Sciences Division, Lawrence Livermore National Laboratory, 7000 East Avenue L-231, Livermore, California 94550, USA. [2]Centre for Star and Planet Formation, University of Copenhagen, Øster Voldgade 5-7 Copenhagen, Denmark. [3]Clermont Université, Université Blaise Pascal, Laboratoire Magmas et Volcans, UMR CNRS 6524, 5 rue Kessler, 63038 Clermont-Ferrand, France. [4]Department of Terrestrial Magnetism, Carnegie Institution, 5241 Broad Branch Road, Northwest, Washington DC 20015-1305, USA.

The Sm–Nd isotopic data define a ^{147}Sm–^{143}Nd isochron age of 4,367 ± 11 Myr (Fig. 2a) that is 73 Myr younger than that determined previously for 60025 (ref. 5). Recalculating the data of ref. 5 using the IsoPlot program (used to calculate the ages reported here) demonstrates that the two ages differ by only 22 Myr. We note that the much larger sample sizes used here allow significantly higher precision as a result of running Nd as a metal instead of an oxide, as well as minimizing concerns about blank correction (see Supplementary Information). These factors, plus the concordance of Sm–Nd and Pb–Pb ages, suggest that the new result improves on the accuracy of earlier work. In addition, we have produced the first lunar ^{146}Sm–^{142}Nd internal isochron, yielding an age of $4,318^{+30}_{-38}$ Myr (Fig. 2b). Although this age is 41–49 Myr younger than the ^{147}Sm–^{143}Nd and Pb–Pb ages, its relatively large uncertainty makes it differ from the other ages by only 8–9 Myr. The minor discordance between chronometers probably reflects the small range in measured ^{142}Nd/^{144}Nd (48 parts per million) and a slight underestimation of the analytical uncertainty on the measured ^{147}Sm/^{144}Nd and/or ^{142}Nd/^{144}Nd ratios. This is illustrated in Fig. 2b by the position of the 4,359 Pb–Pb reference line. Regardless of the cause of the very small discrepancy between Sm–Nd ages, the extremely limited variation in ^{142}Nd/^{144}Nd between the mineral separates is consistent with the young age determined for 60025.

If 60025 represents a flotation cumulate of the lunar magma ocean, its young crystallization age requires the solidification of the magma ocean, and formation of the Moon, to have occurred later than most previous estimates[18–20]. A young Moon is supported by the observation that almost all lunar crustal rocks[18,19], as well as two FANs[8,9], are the same age or younger than 60025. Although model-dependent,

^{142}Nd/^{144}Nd ages for the formation of mare basalt source regions are also young[13,21–23], ranging from $4,313^{+25}_{-30}$ Myr to $4,353^{+21}_{-23}$ Myr, and are concordant with the ^{143}Nd/^{144}Nd model age[24] of 4,360 ± 60 Myr for the last solidification products of the magma ocean (KREEP: potassium, rare-earth elements and phosphate-rich cumulates). Recent ^{182}Hf–^{182}W measurements have shown no radiogenic input to ^{182}W in lunar samples and constrain differentiation of the lunar interior to have occurred after 4.5 billion years (Gyr) ago[25]. The new data for 60025 are thus consistent with formation of the Moon up to about 200 Myr after that of the Solar System.

The oldest terrestrial samples are zircons found in metasedimentary deposits from Jack Hills, Australia, where a single zircon has yielded an age of 4,404 ± 8 Myr (ref. 26). It has REE abundances and an oxygen isotopic composition that indicates it formed from an evolved granitic magma and represents a fragment of continental crust. Taken at face value, the ages of this zircon and 60025 require differentiation on the Moon to occur at least 30 Myr after the differentiation of Earth. This implies that the Moon accreted relatively slowly after the giant impact, or that the Moon retained enough heat to delay cumulate formation.

Alternatively, the young age of 60025 might indicate that it is not a magma ocean product, but rather was produced by a more recent melting event. It should be noted that Sm–Nd ages older than 60025 (Fig. 3) have been reported for three FANs, Y-86032 (4.43 ± 0.03 Gyr; ref. 27), 67075 (4.47 ± 0.07 Gyr; ref. 28) and 67016c (4.53 ± 0.12 Gyr; ref. 10), and one Mg suite sample 15445,17 (4.46 ± 0.07 Gyr; ref. 29). Similarly the oldest lunar zircon, dated at 4,417 ± 6 Myr (ref. 20), comes from the impact melt-breccia 72215 and presumably dates the time of crystallization of a KREEP-rich alkali-suite sample. Finally, some KREEP model ages[18,19,24] are older than 60025, averaging 4,456 ± 65 Myr. Therefore, if 60025 is a flotation cumulate of a magma ocean, these old ages must be in error.

The overlap between the ^{146}Sm–^{142}Nd isotopic systematics of 60025 and those inferred for mare basalt source regions[13,21–23] (Fig. 2b) places additional constraints on FAN petrogenesis. It suggests that the mare basalt sources and FANs formed with very similar initial Nd isotopic compositions at roughly the same time. Admittedly this overlap is not perfect, but is nonetheless remarkable given that the basalt Nd isotopic measurements were completed under different analytical conditions and variably corrected for thermal neutron capture, and the ^{147}Sm/^{144}Nd ratios were calculated using different geochemical models. Geological

Figure 2 | Sm–Nd isochron diagrams. For FAN 60025, plagioclase, pyroxene and whole rocks define a ^{147}Sm–^{143}Nd age of 4,367 ± 11 Myr (**a**) and a ^{146}Sm–^{142}Nd age of $4,318^{+30}_{-38}$ Myr (**b**). **a**, εNd = −0.24 ± 0.09 (MSWD = 0.40). The inset to **a** represents the deviation of individual points from the isochron in epsilon units. **b**, ε^{142}Nd^{CHUR} = +0.10 ±0.04, where CHUR is chondritic uniform reservoir (MSWD = 0.84). ^{146}Sm/^{144}Nd = 0.0085 at 4,568 Myr ago. The inset to **b** represents deviations of lunar whole rocks analysed in refs 13 and 21–23 from the isochron. Uncertainties are ±2σ mean of the population of mass spectrometry ratios.

Figure 3 | Summary of lunar ages. The diagram illustrates the ages of the oldest lunar samples and model ages of cumulate source regions. Symbols refer to ages reported on individual samples. Filled circles represent FANs, open squares represent Alkali and Mg-suite samples, and filled triangles represent model ages. Error bars are uncertainties reported for individual age determinations. Ages to the right of the dashed line are inconsistent with 60025 representing a flotation cumulate of a lunar magma ocean. Data are referenced in the text and the Supplementary Information.

scenarios for the formation of 60025 must therefore include the formation of the low-Ti (pyroxene-rich), high-Ti (ilmenite-rich) and KREEP-rich source regions of the mare basalts. The success of the magma ocean model of lunar differentiation stems largely from its ability to do just this: account for the petrogenesis of all types of lunar source regions by a single mechanism (see, for example, ref. 4). Thus, if 60025 is not a magma ocean cumulate, then another mechanism to re-equilibrate ^{146}Sm–^{142}Nd isotopic systematics of mafic and plagioclase-rich cumulates is required.

One possible mechanism to equilibrate ancient lunar source regions without invoking magma ocean crystallization is for early-formed cumulates to have overturned as a result of density instability of the cumulate pile (for example, ref. 30). In this scenario of serial magmatism, young FANs represent solidification products of plagioclase-rich melts produced during overturn, whereas mare basalts represent later, more mafic, melts of the same cumulate sources. The advantage of this scenario is that it does not require dismissing the handful of older ages determined on lunar samples and it can potentially account for the variable (if somewhat suspect) ages and widespread distribution of FANs on the lunar surface. The disadvantage is that the lunar magma ocean theory is strongly founded upon the petrological, geochemical and isotopic characteristics of FANs such as 60025. Therefore, if 60025 is not a product of the magma ocean, petrologically similar rocks from the FAN suite cannot be used to characterize its initial solidification. Furthermore, the potential for planetary differentiation by magma ocean solidification on all rocky bodies is weakened if the very rocks that led to the development of the magma ocean theory are themselves not its byproduct.

METHODS SUMMARY

Thorough examination of the Apollo sample collection at the Johnson Space Center yielded a 1.88-g pyroxene-rich clast from FAN 60025. It contained 25% mafic minerals (mostly pyroxene) that were separated from plagioclase using a Frantz magnetic separator and hand-picking. Ten milligrams of pyroxene and plagioclase were used to develop washing and sequential dissolution procedures designed to remove Pb contamination from the splits. The 10-mg mineral splits were washed in distilled water, ethanol and acetone, and then in weak hydrobromic acid (HBr). They were then sequentially dissolved using a variety of progressively stronger acids. The effectiveness of washing and sequential dissolution to remove Pb contamination was assessed by determining the Pb isotopic composition of individual washes and dissolutions. These analyses demonstrated that only the pyroxene fraction was suitable for age dating. The washing/sequential dissolution procedure developed for the 10-mg pyroxene split was applied to a larger 105.4-mg pyroxene split. The pyroxene was digested in HF and hydrochloric acid (HCl), spiked with a mixed ^{202}Pb–^{205}Pb tracer, and Pb was purified using HBr-HNO$_3$ based chemistry. Lead was run on a Thermo-Fisher Triton by peak hopping in a secondary electron multiplier. Samarium and neodymium were measured on the remaining pyroxene and plagioclase fractions as well as an additional whole-rock fraction. The samples were digested in HF and nitric acid (HNO$_3$) and spiked with a ^{149}Sm–^{150}Nd tracer prepared at Lawrence Livermore National Laboratory and having 99.988% ^{150}Nd. REEs were separated using HCl and methalactic acid. Total Sm and Nd procedural blanks were 8 pg and 40 pg respectively. The isotopic compositions were determined with a Thermo-Fisher Triton at the Department of Terrestrial Magnetism using double Re filaments. Neodymium data were collected dynamically using 8-s integrations.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Lapen, T. J. et al. A younger age for ALH84001 and its geochemical link to shergottite sources in Mars. Science **328**, 347–351 (2010).
2. Wood, J. A., Dickey, J. S., Marvin, U. B. & Powell, B. N. Lunar anorthosites and a geophysical model of the moon. Proc. Lunar Planet. Sci Conf. **I**, 965 (1970).
3. Smith, J. A. et al. Petrologic history of the Moon inferred from petrography, mineralogy, and petrogenesis of Apollo 11 rocks. Proc. Lunar Planet. Sci Conf. **I**, 1, 149 (1970).
4. Snyder, G. A., Taylor, L. A. & Neal, C. R. A chemical model for generating the source of mare basalts: combined equilibrium and fractional crystallization of the lunar magmasphere. Geochim. Cosmochim. Acta **56**, 3809–3823 (1992).
5. Carlson, R. W. & Lugmair, G. W. The age of ferroan anorthosite 60025: oldest crust on a young Moon? Earth Planet. Sci. Lett. **90**, 119–130 (1988).
6. Hanan, B. B. & Tilton, G. R. 60025—relict of primitive lunar crust. Earth Planet. Sci. Lett. **84**, 15–21 (1987).
7. Norman, M. D., Borg, L. E., Nyquist, L. E. & Bogard, D. D. Chronology, geochemistry, and petrology of a ferroan noritic anorthosite clast from Descartes breccia 67215: clues to the age, origin, structure, and impact history of the lunar crust. Meteorit. Planet. Sci. **38**, 645–661 (2003).
8. Borg, L. E. et al. Isotopic studies of ferroan anorthosite 62236: a young lunar crustal rock from a light rare-earth element-depleted source. Geochim. Cosmochim. Acta **63**, 2679–2691 (1999).
9. Alibert, C., Norman, M. D. & McCulloch, M. T. An ancient age for a ferroan anorthosite clast from lunar breccia 67016. Geochim. Cosmochim. Acta **58**, 2921–2926 (1994).
10. Shirley, D. N. A partially molten magma ocean model. Proc. Lunar Planet. Sci Conf. **XII**, A519–A527 (1983).
11. Warren, P. H. & Wasson, J. T. The origin of KREEP. Rev. Geophys. Space. Phys. **17**, 73–88 (1977).
12. James, O. B., Lindstrom, M. M. & McGee, J. J. Lunar ferroan anorthosites 60025—petrology and chemistry of mafic lithologies. Proc. Lunar Planet. Sci. Conf. **XXI**, 63–87 (1991).
13. Boyet, M. & Carlson, R. W. A highly depleted Moon or a non-magma origin for the lunar crust? Earth Planet. Sci. Lett. **262**, 505–516 (2007).
14. Tera, F. & Wasserburg, G. J. U-Th-Pb systematics in the lunar highland samples from the Luna 20 and Apollo 16 missions. Earth Planet. Sci. Lett. **17**, 36–51 (1972).
15. Nunes, P. D., Knight, R. J., Unruh, D. M. & Tatsumoto, M. The primitive nature of the lunar crust and the problem of initial Pb isotopic compositions of lunar rocks: a Rb-Sr and U-Th-Pb study of Apollo 16 samples. Lunar Planet. Sci. Conf. **V**, 559–561 (1974).
16. Connelly, J. N. & Bizzarro, M. Pb/Pb dating of chondrules from CV chondrites by progressive dissolution. Chem. Geol. **259**, 143–151 (2009).
17. Tera, F. & Wasserburg, G. J. U-Th-Pb systematics of lunar rocks and inferences about lunar evolution. Proc. Lunar Sci. Conf. **V**, 1571–1599 (1974).
18. Nyquist, L. E. & Shih, C.-Yu. The isotopic record of lunar volcanism. Geochim. Cosmochim. Acta **56**, 2213–2234 (1992).
19. Edmunson, J., Borg, L. E., Nyquist, L. E. & Asmeron, Y. A combined Sm-Nd, Rb-Sr, and U-Pb isotopic study of Mg-suite norite 78238: further evidence for early differentiation of the Moon. Geochim. Cosmochim. Acta **73**, 514–527 (2009).
20. Nemchin, A. et al. Timing of crystallization of the lunar magma ocean constrained by the oldest zircon. Nature Geosci. **25**, 133–136 (2009).
21. Nyquist, L. E. et al. ^{146}Sm-^{142}Nd formation interval for the lunar mantle. Geochim. Cosmochim. Acta **59**, 2817–2837 (1995).
22. Rankenburg, K., Brandon, A. D. & Neal, C. R. Neodymium isotope evidence for a chondritic composition of the Moon. Science **312**, 1369–1372 (2006).
23. Brandon, A. D. et al. Re-evaluating ^{142}Nd/^{144}Nd in lunar mare basalts with implications for early evolution and bulk Sm/Nd of the Moon. Geochim. Cosmochim. Acta **73**, 6421–6445 (2009).
24. Carlson, R. W. & Lugmair, G. W. The Sm-Nd history of KREEP. Proc. Lunar Planet. Sci Conf. **IX**, 689–704 (1978).
25. Touboul, M., Kleine, T., Bourdon, B., Palme, H. & Wieler, R. Late formation and prolonged differentiation of the Moon inferred from W isotopes in lunar metals. Nature **450**, 1206–1209 (2007).
26. Wilde, S. A., Valley, J. W., Peck, W. H. & Graham, C. M. Evidence for detrital zircons for the existance of continental crust and oceans on Earth 4.4 Gyr ago. Science **409**, 175–178 (2001).
27. Nyquist, L. et al. Feldspathic clasts in Yamato-86032: remnants of the lunar crust with implications for its formation and impact history. Geochim. Cosmochim. Acta **70**, 5990–6015 (2006).
28. Nyquist, L. E. et al. Lunar crustal history recorded in lunar anorthosites. Lunar Planet. Sci Conf. **XXXXI**, 1383 (2010).
29. Shih, C.-Y. et al. Age of pristine noritic clasts from lunar breccias 15445 and 15455. Geochim. Cosmochim. Acta **57**, 915 (1993).
30. Hess, P. C. & Parmentier, E. M. A model for the thermal and chemical evolution of the Moon's interior; implications for the onset of mare volcanism. Earth Planet. Sci. Lett. **134**, 501–514 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions L.E.B. identified, located and prepared the sample for analysis. J.N.C. performed Pb–Pb measurements. L.E.B., M.B. and R.W.C. completed Sm–Nd measurements. All authors contributed to interpretation of data and preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.E.B. (borg5@llnl.gov).

METHODS

Sample preparation. L.E.B. visited the lunar curation facility at the Johnson Space Center (JSC) in October 2008 and examined several pieces of FAN 60025 searching for a clast that was enriched in mafic minerals. The sample allocated at the JSC weighed 1.88 g and contained approximately 25% mafic minerals consisting of olivine and pyroxene intergrown with large grains of plagioclase. The sample was crushed into coarse fragments, creating some fine powder. Several large fragments were set aside. Mafic and plagioclase fragments were high-graded into two fractions using tweezers. There remained a third fraction that was a 50/50 mixture of mafic minerals and plagioclase, but was too fine-grained to be high-graded using the tweezers. The coarse mafic fraction and the 50/50 mixture were crushed in a sapphire mortar and pestle and sieved yielding fractions of 75–200, 200–325 and <325 mesh. Mineral separations were completed on these fractions using a Frantz isobarrier separator and handpicking under dry conditions to minimize Pb contamination during processing. Pyroxene was also separated from the <325 mesh split of the coarse-grained mafic fraction but was too fine-grained to be handpicked. Visual inspection of the mafic mineral and plagioclase fractions indicated that they were >99% pure.

Pb procedures. The Pb isotopic data were collected in three analytical sessions. The first two sessions were preliminary and used to develop washing and sequential multi-step dissolution procedures to remove contamination from 10-mg splits of plagioclase and pyroxene mineral fractions. The Pb isotopic compositions of individual washes were analysed throughout the washing procedure to determine how much Pb contamination was removed during each washing step, and to assess the next step in the procedure. Once we were convinced that significant quantities of contamination had been removed, sequential digestion of the samples was begun. Preliminary Pb isotopic analysis of wash and sequential digestion fractions of the 10-mg plagioclase split demonstrated that not enough Pb contamination was removed to warrant additional analysis of plagioclase in 60025. The third and final analytical session therefore focused on a 105.4-mg split of the pyroxene fraction. Washing and sequential digestion procedures used in the third session were similar, but not identical, to those used in the preliminary session on the 10-mg pyroxene split. We note that the sequential digestions of the 105.4-mg pyroxene fraction were the only dissolutions designed to define the age of the sample. Nevertheless, the isotopic compositions of all of the preliminary washes and digestions of the plagioclase and both pyroxene splits are reported in the Supplementary Information.

The fractions for all sessions were pre-washed in five cycles of distilled water, ethanol and acetone before progressive weak-acid washing. The Pb isotopic composition of these washes was not determined. Next, the 10-mg splits were washed in water followed by six washes in very weak HBr. The final wash was in 0.5 M HBr. After washing, the 10-mg plagioclase and pyroxene splits were sequentially dissolved in a multi-step procedure that used progressively stronger acids. Despite the small amount of Pb in the final dissolution steps of the 10-mg pyroxene split, the analyses of the washes and sequential dissolutions of this fraction showed two encouraging features for geochronology: (1) as cleaning progressed, the Pb from the residual material became more radiogenic, and (2) the analyses of the sequential dissolutions defined an approximately linear trend despite a significant correction for Pb blank on most analyses.

From this information it was clear that the pyroxene required extensive cleaning in water, acetone, ethanol and weak acids, with a final cleaning step using 0.5 M HBr before starting the multi-step dissolution. The dissolution procedure successively used 6 M HCl, 1 M HF, 7 M HF, 7 M HNO$_3$, and 28 M HF + 14 M HNO$_3$ to produce the Px-2 (L1)–(L5) fractions, respectively. Lead was separated from matrix elements by passing the samples twice through a HBr–HNO$_3$-based chemical separation procedure using 0.055 ml of Eichrom anion resin[16]. The purified Pb was analysed on a Thermo-Fisher Triton thermal ionization mass spectrometer equipped with nine Faraday detectors and one axial secondary-electron-multiplier ion-counting system. Lead was loaded onto previously outgassed zone-refined Re filaments with silica gel made from silicic acid following ref. 31. All analyses were made by sequentially peak jumping the ion beams into the central secondary-electron-multiplier ion-counting system. Samples were spiked with a ^{202}Pb–^{205}Pb double spike to allow for an internal correction for instrument mass fractionation. Samples are corrected for a Pb blank added during the chemical separation according to replicate blanks run during the same session. Lead blanks from the chemistry during this work ranged from 0.25 to 0.50 pg. Samples are also corrected for a 0.2-pg Pb loading blank. The data with 2σ errors are presented in the Supplementary Information.

Sm–Nd procedures. Samples for Sm–Nd analyses were derived from the aliquots of the plagioclase and whole-rock fractions used for Pb–Pb analyses, plus an additional whole-rock fraction. The samples were dissolved in 3:1 mixtures of concentrated HF:HNO$_3$ in sealed Savillex beakers on a hot plate at 90 °C. Before dissolution, the samples were spiked with ^{149}Sm and ^{150}Nd, for concentration

determination. A new mixed spike consisting of 99.988% purity ^{150}Nd and 97.7% ^{149}Sm was prepared at Lawrence Livermore National Laboratory for this work and calibrated against the AMES metal Sm–Nd mixed standard described in ref. 32. After three days on the hot plate, the samples were evaporated to dryness, treated twice with concentrated HNO$_3$, drying in between treatments, and then dissolved in 6 M HCl. The volume of HCl was increased until a clear solution was obtained. Once complete dissolution was achieved, the samples were dried and then redissolved in 5 ml of 1 M HCl + 0.1 M HF. The samples were centrifuged to remove any undissolved material. The clear liquid was loaded on 20-cm-long, 1-cm-diameter quartz columns filled with AG50W-X8 resin. High-field-strength elements were eluted in the sample loading solution and an additional 5 ml of 1 M HCl/0.1 M HF. Major elements and large ion lithophile trace elements were then eluted in 2.5 M HCl with heavy-REE and light-REE splits eluted in 4 M HCl. The light-REE split was processed further to separate Sm and Nd using the procedure described most recently in ref. 33. The column procedure used here has a total blank of 12 pg Nd and 2.3 pg Sm. The sample data were corrected using blanks of 40 pg Nd and 8 pg Sm to account for the volume of acid used for dissolution and because each sample was split between two columns for the first step of the chemical separation. The smallest Sm–Nd amounts analysed here were the 89 ng of Nd obtained from the mafic (olivine + pyroxene) mineral split and the 40 ng of Sm extracted from the plagioclase. The blanks listed above constitute less than 0.05% of the analysed Sm and Nd and are negligible. Concentration data are presented in the Supplementary Information.

Neodymium was loaded in 3 M HCl onto a Re filament and analysed as Nd$^+$ using a second Re filament for ionization. Isotope ratios were measured with the Department of Terrestrial Magnetism's Thermo-Fisher Triton thermal ionization mass spectrometer using a two-mass-step procedure with ^{145}Nd and then ^{143}Nd in the axial Faraday detector. This procedure calculates ^{142}Nd/^{144}Nd dynamically to eliminate Faraday detector biases. All other Nd-isotope ratios are measured statically along with potential interfering species Ce and Sm. Each step integrates the signal for 8 s, leading to 16 s of signal integration for each ratio except for ^{142}Nd/^{144}Nd and ^{150}Nd/^{144}Nd that have 8 s of signal integration per ratio. For the plagioclase, 480 ratios were obtained at an average ^{142}Nd = 1.2×10^{-11} A. The pyroxene produced 420 ratios at an average ^{142}Nd = 1.6×10^{-11} A. The whole-rock analysis lasted only for 130 ratios at an average ^{142}Nd = 8.5×10^{-12} A. Errors reported for the sample measurements are the 2σ-mean from the individual mass spectrometer runs. All data are corrected ratio by ratio during mass spectrometry for Ce interference (using ^{142}Ce/^{140}Ce = 0.125) and Sm interference (using the measured Sm isotopic composition of each sample). Six 540 ratio runs of the JNdi Nd standard at signal sizes of $2–5 \times 10^{-11}$ A were run in the same barrel as the sample analyses. The results of these standard runs are listed in the Supplementary Information along with the isotopic compositions measured for each sample. Because the samples were total-spiked, the ^{150}Nd/^{144}Nd and ^{149}Sm/^{152}Sm ratios are considerably higher than standard Nd and Sm. All Nd ratios except for ^{146}Nd/^{144}Nd are corrected for mass fractionation to ^{146}Nd/^{144}Nd = 0.7219, using exponential mass dependency. All but ^{146}Nd/^{144}Nd and ^{150}Nd/^{144}Nd are corrected for the small spike contribution to these isotopes from the highly enriched ^{150}Nd spike. The ^{146}Nd/^{144}Nd ratio reported in the Supplementary Information is the measured value before fractionation and spike correction.

Samarium was loaded in 2 M HNO$_3$ onto a Re filament and analysed as Sm$^+$ using a second Re filament for ionization. All Sm isotopes along with potential interference from Nd (^{146}Nd) were measured statically for 180 ratios of 8-s integration each. Corrections for Nd interference on Sm were made using the measured Nd isotopic composition of each sample. Sm concentrations were calculated assuming normal Sm as previous measurements of unspiked Sm from 60025 show no resolvable deficit in ^{149}Sm that could be attributable to neutron capture[13,24]. Both the whole-rock and mafic fractions measured here have spike-corrected ^{150}Sm/^{152}Sm ratios only marginally higher than normal Sm, in agreement with previous unspiked Sm measurements[13,24] in different samples of 60025. The Sm isotopic compositions are corrected for fractionation to ^{147}Sm/^{152}Sm = 0.56081 with the exception of the ^{147}Sm/^{152}Sm ratio, which is reported as measured in the Supplementary Information. All ratios except ^{149}Sm/^{152}Sm and ^{147}Sm/^{152}Sm reported in the Supplementary Information have also been corrected for the contribution from the ^{149}Sm spike.

31. Gerstenberger, H. & Haase, G. A highly effective emitter substance for mass spectrometric Pb isotope ratio determinations. *Chem. Geol.* **136,** 309–312 (1997).
32. Boyet, M. & Carlson, R. W. ^{142}Nd evidence for early (>4.53 Ga) global differentiation of the silicate Earth. *Science* **309,** 576–581 (2005).
33. Boyet, M., Carlson, R. W. & Horan, M. Old Sm-Nd ages for cumulate eucrites and redetermination of the solar system initial ^{146}Sm/^{144}Sm ratio. *Earth Planet. Sci. Lett.* **291,** 172–181 (2010).

LETTER

Real-time quantum feedback prepares and stabilizes photon number states

Clément Sayrin[1], Igor Dotsenko[1], Xingxing Zhou[1], Bruno Peaudecerf[1], Théo Rybarczyk[1], Sébastien Gleyzes[1], Pierre Rouchon[2], Mazyar Mirrahimi[3], Hadis Amini[2], Michel Brune[1], Jean-Michel Raimond[1] & Serge Haroche[1,4]

Feedback loops are central to most classical control procedures. A controller compares the signal measured by a sensor (system output) with the target value or set-point. It then adjusts an actuator (system input) to stabilize the signal around the target value. Generalizing this scheme to stabilize a micro-system's quantum state relies on quantum feedback[1–3], which must overcome a fundamental difficulty: the sensor measurements cause a random back-action on the system. An optimal compromise uses weak measurements[4,5], providing partial information with minimal perturbation. The controller should include the effect of this perturbation in the computation of the actuator's operation, which brings the incrementally perturbed state closer to the target. Although some aspects of this scenario have been experimentally demonstrated for the control of quantum[6–9] or classical[10,11] micro-system variables, continuous feedback loop operations that permanently stabilize quantum systems around a target state have not yet been realized. Here we have implemented such a real-time stabilizing quantum feedback scheme[12] following a method inspired by ref. 13. It prepares on demand photon number states (Fock states) of a microwave field in a superconducting cavity, and subsequently reverses the effects of decoherence-induced field quantum jumps[14–16]. The sensor is a beam of atoms crossing the cavity, which repeatedly performs weak quantum non-demolition measurements of the photon number[14]. The controller is implemented in a real-time computer commanding the actuator, which injects adjusted small classical fields into the cavity between measurements. The microwave field is a quantum oscillator usable as a quantum memory[17] or as a quantum bus swapping information between atoms[18]. Our experiment demonstrates that active control can generate non-classical states of this oscillator and combat their decoherence[15,16], and is a significant step towards the implementation of complex quantum information operations.

A Fock state with n photons is hard to generate and very fragile. Prepared in a cavity of damping time T_c, it survives on average for T_c/n before undergoing a quantum jump towards the $|n - 1\rangle$ Fock state. In contrast, classical Glauber states[19], which are coherent superpositions of Fock states with an average photon number $\bar{n}$ and a Poisson photon number probability distribution $P(n) = \exp(-\bar{n})\,(\bar{n}^n/n!)$, are much easier to prepare and more robust. Glauber states are easily obtained by coupling the initially empty cavity to a classical field source for a fixed amount of time. This operation amounts to the translation of the field in its phase space from the vacuum ($\bar{n} = 0$ coherent state) to a final coherent state having an amplitude $\alpha = \sqrt{\bar{n}_0}$ with a mean photon number $\bar{n}_0$. After the source is switched off, the field remains in a coherent state with an exponentially decaying amplitude, $\bar{n}$ becoming $\bar{n}(t) = \bar{n}_0\exp(-t/T_c)$.

Experimental methods to prepare Fock states in a cavity C start from a coherent state and exploit the coupling of the field to two-level qubits[14,20,21]. A deterministic procedure feeds quanta one at a time into the field initially in vacuum by swapping its energy with a qubit periodically re-pumped into its excited state[21]. This method, which has

been generalized to synthesize arbitrary superpositions of Fock states[22], cannot counteract decoherence because it does not provide real time information on the actual field state in C. Fock states can also be prepared by a quantum non-demolition (QND) measurement performed on an initial coherent state with $\bar{n}_0 \neq 0$ (ref. 14). Atomic qubits probe the field one at a time and the photon number is progressively pinned down to an inherently random value, the probability $P(n)$ for finding n being the value corresponding to the initial coherent field. This QND method provides real time information about the field state history during the process. This information can be used for a deterministic steering of the field towards a target Fock state $|n_t\rangle$, as well as for detection and subsequent correction of quantum jump events. We have performed a quantum feedback experiment by combining the detection of successive atoms with field phase-space translations of controlled amplitudes. We thus prepare Fock states $|n_t\rangle$ on demand and, on average, stabilize them by bringing the field back into them after decoherence-induced quantum jumps.

The experiment is performed in a superconducting cavity C with $T_c = 65$ ms cooled to 0.8 K (see Fig. 1 and Supplementary Methods). It is initially fed by the source S which prepares a coherent state with a real amplitude $\alpha_t = \sqrt{n_t}$. The quantum sensors are circular Rydberg atoms prepared in B at regular time intervals ($T_a = 82$ μs)[18,23]. The number of Rydberg atoms in each sample obeys Poisson statistics, with 0.6 atoms per sample on average. The atomic states $|g\rangle$ and $|e\rangle$ with principal quantum numbers 50 and 51 are the 0 and 1 states of a qubit slightly off-resonant with cavity C (atom–cavity detuning $\delta/2\pi = 245$ kHz). The qubit coherence undergoes in C a light-induced phase shift linear in the photon number (phase-shift per photon $\phi_0 = 0.256\pi$). This phase shift is measured by a Ramsey interferometer (R_1 and R_2). Detecting each atomic sample in D provides partial information about the number of photons in C.

Each iteration of the feedback loop[12] consists of a sample detection by the detector D, a cavity field state estimation by the controller K and a field translation performed by the actuator S (Fig. 1). In each iteration, K first updates its estimation of the field density operator ρ on the basis of the detection outcome, and corrects this estimation by taking into account the effect of cavity relaxation at finite temperature during the iteration time T_a. It then computes the amplitude α of the translation described by the operator $D(\alpha) = \exp(\alpha a^\dagger - \alpha^* a)$ (here a is the photon annihilation operator, $\dagger$ and $*$ denote Hermitian and complex conjugate, respectively). Because the initial and target density operators are real, we restrict the translations to real values of α. The field translation minimizes a proper 'distance' $d(\rho_t, D(\alpha)\rho D(-\alpha))$ (defined below) between the displaced state and the target state $\rho_t = |n_t\rangle\langle n_t|$. Finally, at the end of each feedback loop iteration, K calculates the translated field's state, which is to be used at the beginning of the next iteration. Note that this quantum state estimation, performed on a single quantum trajectory, cannot be obtained from the measurement data only. It also incorporates all available information on the state preparation, displacements and relaxation.

[1]Laboratoire Kastler Brossel, ENS, UPMC–Paris 6, CNRS, 24 rue Lhomond, 75005 Paris, France. [2]Centre Automatique et Systèmes, Mathématiques et Systèmes, Mines ParisTech, 60 Boulevard Saint-Michel, 75272 Paris Cedex 6, France. [3]INRIA Paris-Rocquencourt, Domaine de Voluceau, BP 105, 78153 Le Chesnay Cedex, France. [4]Collège de France, 11 place Marcelin Berthelot, 75231 Paris Cedex 05, France.
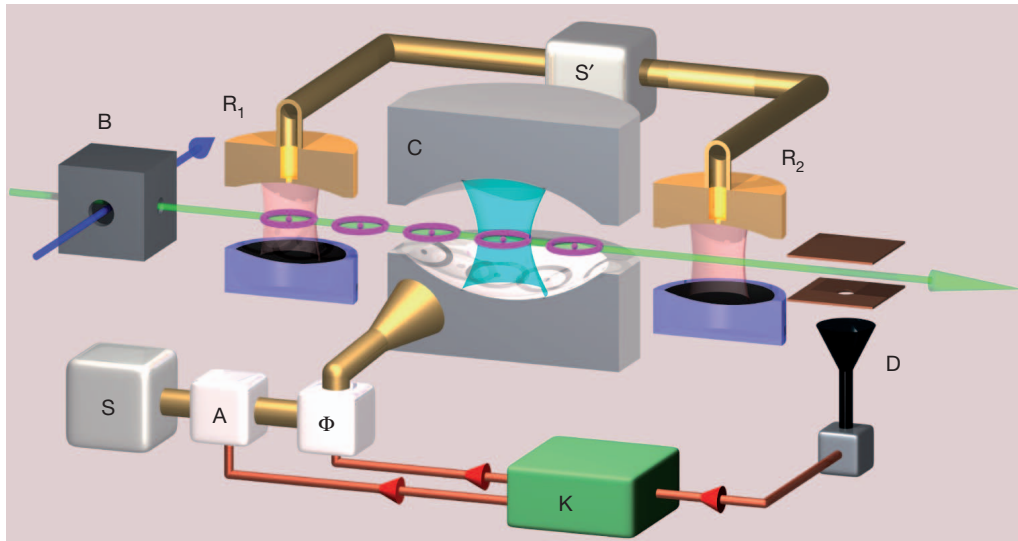
Figure 1 | Scheme of the quantum feedback set-up. An atomic Ramsey interferometer (auxiliary cavities R_1 and R_2) sandwiches the superconducting Fabry–Perot cavity C resonant at 51 GHz and cooled to 0.8 K (the mean number of blackbody photons is 0.05). The pulsed classical source S′ induces $\pi/2$ pulses resonant with the $|g\rangle \rightarrow |e\rangle$ transition in R_1 and R_2 (with relative phase ϕ_r) on the velocity-selected (250 m s^{-1}) Rydberg atom qubits (purple circles) prepared by laser excitation (blue arrow) from a rubidium atomic beam (green arrow) in B. The field-ionization detector D measures the qubits in the e/g basis with a 35% detection efficiency and an error rate of a few per cent (Supplementary Methods). The actuator S feeds cavity C by diffraction on the mirror edges. The controller K (a CPU-based ADwin Pro-II system) collects information from D to determine the real translation amplitude α applied by S. It sets the S-pulse duration through a PIN diode switch A (63-μs pulse for $|\alpha| = 0.1$) as well as a 180° phase-shifter Φ controlling the sign of α.
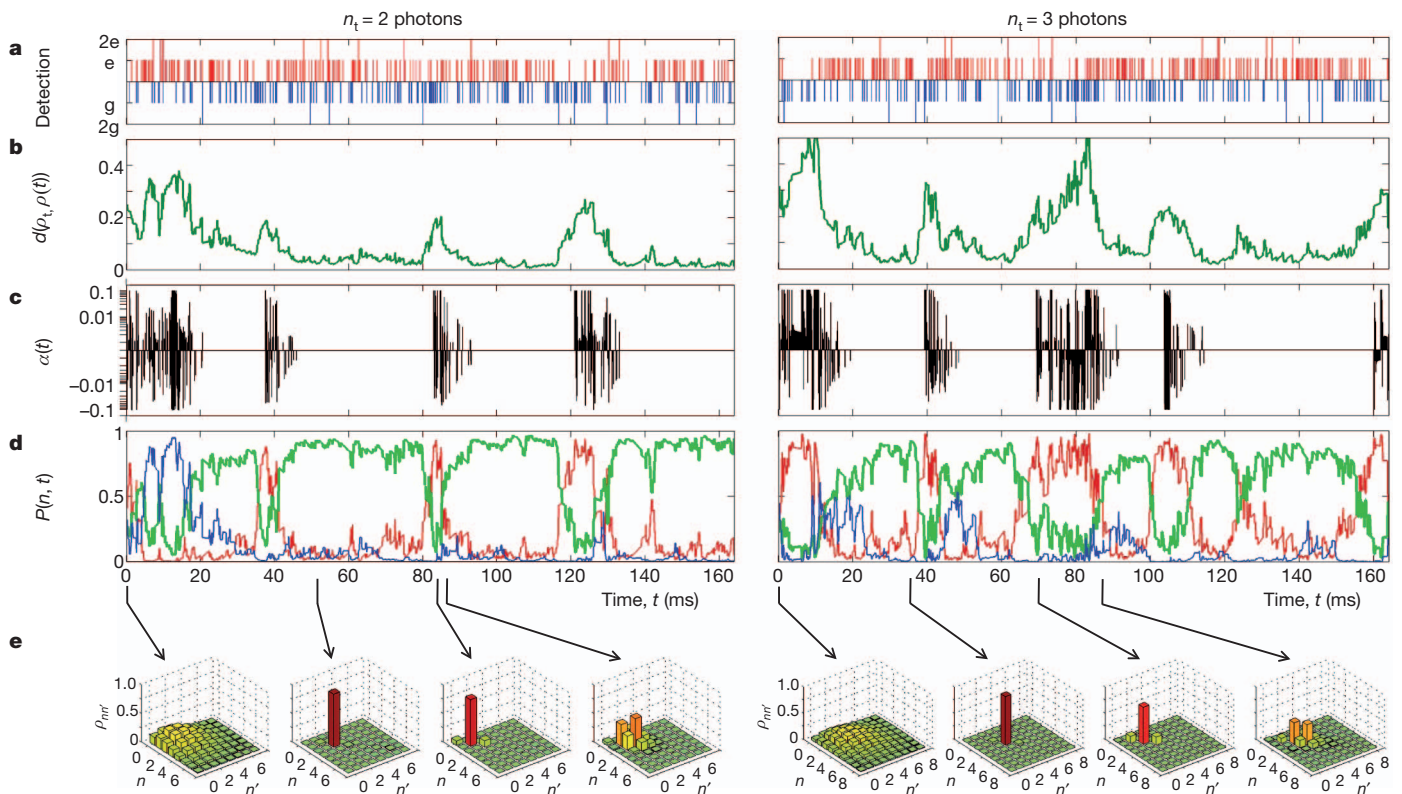
Figure 2 | Individual quantum feedback trajectories. Two feedback runs lasting 164 ms (2,000 loop iterations) stabilizing $|n_t = 2\rangle$ (left column) and $|n_t = 3\rangle$ (right column). The phase-shift per photon, $\phi_0 = 0.256\pi$, allows controller K to discriminate n values between 0 and 7. For $n_t = 2$, the Ramsey phase is $\phi_r = -0.44$ rad, corresponding to nearly equal e and g detection probabilities when $n = 2$. For $n_t = 3$, two Ramsey phases $\phi_{r,1} = -0.44$ rad and $\phi_{r,2} = -1.24$ rad are alternatively used, corresponding to equal e and g probabilities when $n = 2$ and $n = 3$, respectively. **a**, Sequences of qubit detection outcomes. The detection results are shown as blue downward bars for g and red upward bars for e. Two-atom detections in the same state appear as double-length bars. **b**, Estimated distance between the target and the actual state. **c**, Applied α-corrections (shown on a log scale as $\mathrm{sgn}(\alpha)\log|\alpha|$). **d**, Photon number probabilities estimated by K: $P(n = n_t)$ is in green, $P(n < n_t)$ in red, $P(n > n_t)$ in blue. **e**, Field density operators ρ in the Fock-state basis estimated by K at four different times marked by arrows.
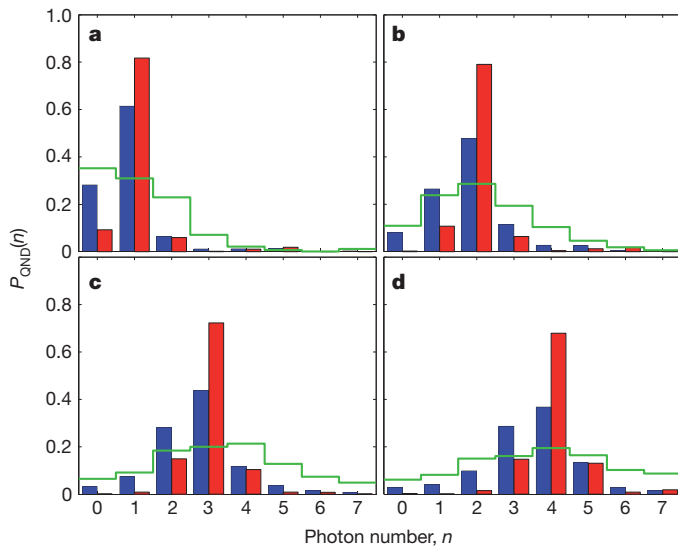
Figure 3 | Photon number histograms following quantum feedback iterations. Plots **a**, **b**, **c** and **d** correspond to the target photon number states $n_t = 1$, 2, 3 and 4, respectively. The red histograms correspond to about 3,900 trajectories stopped when $P(n_t)$ has reached the threshold value (0.8) for three successive iterations. These histograms describe the field at the time when the controller K has certified the 'success' of the quantum feedback procedure. The blue histograms correspond to 4,000 trajectories stopped at a fixed time (164 ms) and describe the feedback procedure steady-state. These histograms are reconstructed by a method independent of the feedback estimator. After interrupting the feedback, we record ten additional QND qubit samples (~2 detected atoms) with a Ramsey interferometer phase ϕ_r chosen in sequence among 4 values ($\phi_r = 1.17$, 0.36, -0.44 and -1.24 rad). From these additional qubit detections, we reconstruct the final $P_{QND}(n)$ distribution for each ensemble of trajectories by a maximum likelihood algorithm. Statistical error of the reconstructed $P_{QND}(n)$ for different target states is about 0.01–0.02 for $n = n_t$ and $n_t \pm 1$, and it is significantly smaller than 0.01 for other photon numbers (see Supplementary Methods). The green histograms give the initial coherent state photon number distributions (a similar reconstruction was performed with a fixed time stop immediately after initial field injection).

In an ideal experiment, with exactly one atom prepared and perfectly detected in each sample, a detection in $|e\rangle$ or $|g\rangle$ would actualize the state estimation by the mapping $\rho \rightarrow M_j \rho M_j^\dagger / \mathrm{Tr}(\rho M_j^\dagger M_j)$ $(j = e, g)$, with $M_e = \cos[(\phi_r + \phi_0(N + \frac{1}{2}))/2]$ and $M_g = \sin[(\phi_r + \phi_0(N + \frac{1}{2}))/2]$ where ϕ_r is the tunable phase of the Ramsey interferometer and $N = a^\dagger a$ the photon number operator. This qubit detection is a weak measurement of N associated with the positive operator valued measure (POVM), $\Pi_j = M_j^\dagger M_j$. In the actual experiment, the measurement-induced state mapping takes into account all known and independently measured imperfections: the possibility of 0 and 2 atoms in atomic samples, finite detection efficiency and wrong atomic state assignment (see Supplementary Methods for details). If, for instance, no detection occurs, there is a probability that no atom was present in the sample, in which case the field state does not change. There is another probability that the detector has failed to detect a single qubit, in which case the field should be updated according to the mapping $\rho \rightarrow \sum_j M_j \rho M_j^\dagger$. It is also possible that the detector has missed two qubits, in which case the updating would be $\rho \rightarrow \sum_{jj'} M_{j'} M_j \rho M_j^\dagger M_{j'}^\dagger$ $(j, j' = e, g)$. The probabilities that these situations have occurred, conditioned to the fact that no detection was made, are obtained by a classical Bayesian inference argument. Similar Bayesian reasonings are used to infer the probabilities which affect the mapping when one or two qubits are detected. The state estimation also takes into account the back-action on the field of the yet undetected samples which are on their 344-µs-long flight from C to D (Fig. 1).

The control law relies on a Lyapunov-based state stabilization[24]. Its efficiency depends upon the definition of the distance $d(\rho_t, \rho)$ (the
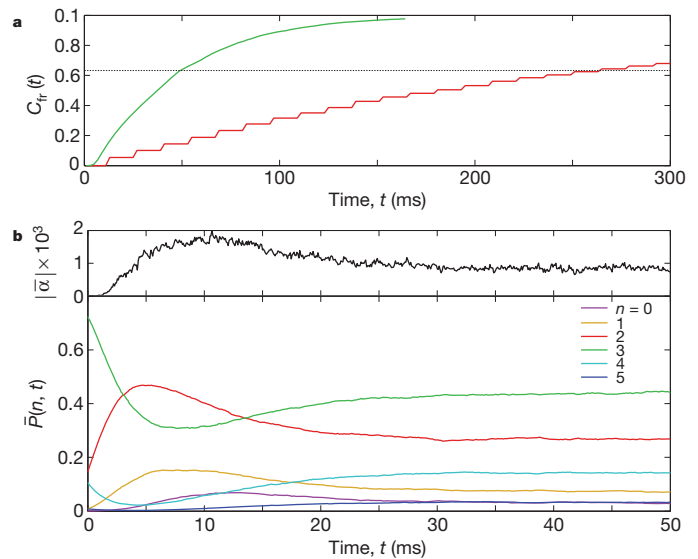
Figure 4 | Performance of the quantum feedback procedure. a, Time evolution of the fraction of individual field trajectories, $C_{fr}(t)$, that have reached the fidelity threshold (0.8) while converging towards $|n_t = 3\rangle$ in quantum feedback sequences (smooth line) and in passive QND 'trials' (stepped line). Statistics performed over 4,000 and 2,131 trajectories, respectively. The same Ramsey phase settings as in Fig. 2 have been used for both feedback and QND sequences. **b**, Recovery from a quantum jump: the lower plot shows probabilities $\bar{P}(n, t)$ estimated by K and averaged over 2,561 trajectories, following the preparation at $t = 0$ of the Fock state $|n = 2\rangle$ by a QND measurement of an initial coherent state (colour code shown at right for the different values of n). The Ramsey phase settings are the same as in Fig. 2 for $n_t = 3$. The initial field density matrix of the field estimation algorithm is diagonal and corresponds to the red histogram in Fig. 3c. The experiment thus simulates the reaction of the quantum feedback procedure to a $|3\rangle \rightarrow |2\rangle$ quantum jump occurring at $t = 0$, after the field has converged to the target. The upper plot in **b** shows the variation of the average modulus of the injection amplitude $|\bar{\alpha}(t)|$. Initially zero, $|\bar{\alpha}|$ grows rapidly to a maximum while the quantum jump is reversed. The controller finally quiets and $|\bar{\alpha}|$ returns to its average steady-state value.

control Lyapunov function) between the field estimation ρ and the target $\rho_t = |n_t\rangle\langle n_t|$. In the simulations described in ref. 12, the simple definition $d = 1 - \langle n_t|\rho|n_t\rangle$ was used. This distance vanishes when the target is reached, but it does not discriminate the $n \neq n_t$ Fock states whose distances to the target are all equal to 1. A better choice defines the distance as $d = 1 - \mathrm{Tr}(\Lambda^{(n_t)}\rho)$, where $\Lambda^{(n_t)}$ is a diagonal matrix with $\langle n_t|\Lambda^{(n_t)}|n_t\rangle = 1$ and the other elements $\langle n|\Lambda^{(n_t)}|n\rangle$ $(n \neq n_t)$ decreasing monotonically with $|n - n_t|$. In this case, d carries information not only about the probability that the field contains n_t photons, but also about how far from n_t non-negligible $P(n)$ values are found. The $\Lambda^{(n_t)}$ matrix is optimized by performing simulations of feedback trajectories and adjusting the $\Lambda_{nn}^{(n_t)}$ coefficients to obtain the fastest convergence. Based on this value of $\Lambda^{(n_t)}$, K searches, at each iteration step, for the α value which minimizes $d(\rho_t, D(\alpha)\rho D(-\alpha))$. To reduce the computation time, it uses an expansion of $D(\alpha)$ up to second order and determines, under this approximation, an optimal field translation with α in the $[-0.1, +0.1]$ interval (see Supplementary Methods).

Figure 2 shows the experimental records of two 164-ms-long feedback sequences aiming at $|n_t = 2\rangle$ (left column) and $|n_t = 3\rangle$ (right column). The measurement outcomes (Fig. 2a) are fed into K, which updates the distance to the target (Fig. 2b) and computes the optimal field translation applied by S (Fig. 2c). This results in the estimated probabilities for finding $n = n_t$, $n < n_t$ and $n > n_t$ number of photons in C (Fig. 2d). After an initial transient period lasting about 20 ms (240 iterations, about 50 detected atoms), the distance to the target drops to a small value and the field reaches $|n_t\rangle$ with a fidelity $\langle n_t|\rho|n_t\rangle \approx 0.8$. The actuator operates during the convergence phase and then quiets down until the field undergoes a quantum jump towards $|n_t - 1\rangle$. The

distance to the target then features a sudden burst, inducing S to become active again, until the target state is restored, in a time of about 10–20 ms (120–240 iterations). Later quantum jumps are corrected in the same way. The rate of quantum jumps increases with n, which explains why S is somewhat more active for $n_t = 3$ than for $n_t = 2$, with a slightly reduced average fidelity. Similar recordings obtained for $n_t = 1$ and 4 are shown in Supplementary Methods.

Figure 2e shows snapshots of the density operator ρ as estimated by the feedback controller K. For each sequence, we have represented from left to right the initial coherent state, the states after the convergence has been observed, shortly after a quantum jump has been detected, and finally during the recovery from the jump. Note that the initially large off-diagonal elements $\rho_{nn'}(n \neq n')$ vanish when the field state reaches the target represented by a single peaked diagonal matrix. A quantum jump is detected as a fast increase of the $|n_t - 1\rangle$ state probability at the expense of that of the $|n_t\rangle$ state, without build-up of off-diagonal elements. The recovery from the jump is due to small coherent field injections which create transient $\rho_{nn'}$ coherences between Fock states close to $n = n_t$. Supplementary Movies 1 and 2 show the complete evolution of the field density operator during the feedback loops displayed in Fig. 2.

For each n_t value, we have recorded large sets of feedback trajectories with two different stopping conditions: 4,000 of them are stopped by the controller at 164 ms, as in Fig. 2 (fixed time stop), and about 3,900 are stopped when $P(n_t)$ is found by K to be greater than 0.8 in 3 successive iterations (fixed fidelity stop). For each n_t and stopping condition, the final ensemble-averaged photon number distribution $P_{QND}(n)$ is reconstructed independently from the K estimation, using additional probe atoms sent immediately after the interruption of the feedback loop (see Supplementary Methods). The blue and red bars in Fig. 3 give the values of $P_{QND}(n)$ obtained for the fixed time stop and the fixed fidelity stop, respectively, for $n_t = 1$ to 4. For reference, the green histograms show the measured photon number distribution of the initial coherent state, well described by the Poisson statistics. The high values of the red bars peaking at n_t show the actual fidelity of the state preparation. The blue bar histograms are somewhat broader than the red ones because, on average, the field resides for fractions of time in states with $n \neq n_t$ due to the finite time it takes to correct a quantum jump. These 'fixed time stop' histograms are however narrower than the initial ones of the coherent field, with $P_{QND}(n_t)$ about 2 times larger than the corresponding value for the coherent state.

We have also analysed the speed of convergence towards the target. Figure 4a shows the fraction of trajectories that have reached the fidelity threshold (0.8) for $n_t = 3$ as a function of time. The convergence time (at which 63% of the trajectories have converged) is 50 ms. We compare this result with that of an optimized trial-and-error projection method based on a QND measurement. The photon number of an initial Glauber state with amplitude $\sqrt{n_t}$ is measured by QND probe qubits sent for a fixed time τ. The preparation is declared successful if the inferred probability for n_t is >0.8. Otherwise, the field is reset to the initial state and the procedure repeated until the threshold is reached. Choosing $\tau = 14$ ms optimizes the convergence rate. The stepped line in Fig. 4a shows that the convergence time is now 250 ms, 5 times longer than that of the quantum feedback method.

Finally, we have investigated the dynamics of recovery from a quantum jump out of $|n_t = 3\rangle$. We prepare the field in the $|n_t - 1 = 2\rangle$ Fock state, using a projective QND measurement. We then start a feedback loop with the initial estimated photon number distribution given by the red histogram in Fig. 3c. We thus simulate experimentally the situation in which the field has suddenly jumped into $|n = n_t - 1\rangle$ while K still 'believes' that $n = n_t$. Figure 4b presents the time evolution of the subsequent values of the photon number distribution $\bar{P}(n,t)$ estimated by K and averaged over 2,561 trajectories. Within about 3 ms (~7 detected atoms), K 'realizes' that the jump has occurred (shown by the rapid drop of $\bar{P}(n_t,t)$ and the fast rise of $\bar{P}(n_t - 1,t)$) and activates the control injection. The field comes back to

its steady state (with $\bar{P}(n_t) = 0.43$, this value being limited by subsequent random quantum jumps) within ~15 ms.

We have implemented a real-time quantum feedback procedure that generates photon number states on demand and stabilizes them by reversing the effects of decoherence-induced quantum jumps. This experiment, which combines quantum measurements and deterministic corrections, shows obvious similarities with quantum error correction codes[25] demonstrated with photons[26], ions[27], spins[28] or superconducting qubits[29]. The long cavity damping time of our cavity quantum electrodynamics set-up is an asset, because it allows the controller to perform in real time complex estimation and optimization operations. We plan to perform a variant of this experiment in which the classical actuator source will be replaced by Rydberg atoms delivering single photons into the cavity. The same set-up could also be used to perform adaptive photon number measurements in which the successive qubit settings will be modified in real time, taking into account the results of previous detections[23]. We are also considering applying similar quantum feedback strategies to the stabilization of even more exotic states, such as 'Schrödinger cat' states of radiation[30].

1. Wiseman, H. M. Quantum theory of continuous feedback. *Phys. Rev. A* **49**, 2133–2150 (1994).
2. Doherty, A. C., Habib, S., Jacobs, K., Mabuchi, H. & Tan, S. M. Quantum feedback control and classical control theory. *Phys. Rev. A* **62**, 012105 (2000).
3. Wiseman, H. M. & Milburn, G. J. *Quantum Measurement and Control* (Cambridge Univ. Press, 2009).
4. Aharonov, Y. & Vaidman, L. Properties of a quantum system during the time interval between two measurements. *Phys. Rev. A* **41**, 11–20 (1990).
5. Peres, A. & Wootters, W. K. Optimal detection of quantum information. *Phys. Rev. Lett.* **66**, 1119–1122 (1991).
6. Nelson, R. J., Weinstein, Y., Cory, D. & Lloyd, S. Experimental demonstration of fully coherent quantum feedback. *Phys. Rev. Lett.* **85**, 3045–3048 (2000).
7. Smith, W. P., Reiner, J. E., Orozco, L. A., Kuhr, S. & Wiseman, H. M. Capture and release of a conditional state of a cavity QED system by quantum feedback. *Phys. Rev. Lett.* **89**, 133601 (2002).
8. Cook, R. L., Martin, P. J. & Geremia, J. M. Optical coherent state discrimination using a closed-loop quantum measurement. *Nature* **446**, 774–777 (2007).
9. Gillett, G. G. et al. Experimental feedback control of quantum systems using weak measurements. *Phys. Rev. Lett.* **104**, 080503 (2010).
10. Bushev, P. et al. Feedback cooling of a single trapped ion. *Phys. Rev. Lett.* **96**, 043003 (2006).
11. Kubanek, A. et al. Photon-by-photon feedback control of a single-atom trajectory. *Nature* **462**, 898–901 (2009).
12. Dotsenko, I. et al. Quantum feedback by discrete quantum nondemolition measurements: Towards on-demand generation of photon-number states. *Phys. Rev. A* **80**, 013805 (2009).
13. Geremia, J. M. Deterministic and nondestructively verifiable preparation of photon number states. *Phys. Rev. Lett.* **97**, 073601 (2006).
14. Guerlin, C. et al. Progressive field-state collapse and quantum non-demolition photon counting. *Nature* **448**, 889–893 (2007).
15. Brune, M. et al. Process tomography of field damping and measurement of Fock state lifetimes by quantum nondemolition photon counting in a cavity. *Phys. Rev. Lett.* **101**, 240402 (2008).
16. Wang, H. et al. Measurement of the decay of Fock states in a superconducting quantum circuit. *Phys. Rev. Lett.* **101**, 240401 (2008).
17. Maître, X. et al. Quantum memory with a single photon in a cavity. *Phys. Rev. Lett.* **79**, 769–772 (1997).
18. Raimond, J. M., Brune, M. & Haroche, S. Manipulating quantum entanglement with atoms and photons in a cavity. *Rev. Mod. Phys.* **73**, 565–582 (2001).
19. Glauber, R. J. Coherent and incoherent states of the radiation field. *Phys. Rev.* **131**, 2766–2788 (1963).
20. Varcoe, B. T. H., Brattke, S., Weidinger, M. & Walther, H. Preparing pure photon number states of the radiation field. *Nature* **403**, 743–746 (2000).
21. Hofheinz, M. et al. Generation of Fock states in a superconducting quantum circuit. *Nature* **454**, 310–314 (2008).
22. Hofheinz, M. et al. Synthesizing arbitrary quantum states in a superconducting resonator. *Nature* **459**, 546–549 (2009).
23. Haroche, S. & Raimond, J. M. *Exploring the Quantum: Atoms, Cavities and Photons* (Oxford Univ. Press, 2006).
24. Khalil, H. K. *Nonlinear Systems* (Prentice Hall, 2001).
25. Steane, A. M. Error correcting codes in quantum theory. *Phys. Rev. Lett.* **77**, 793–797 (1996).
26. Lu, C.-Y. et al. Experimental quantum coding against qubit loss error. *Proc. Natl Acad. Sci. USA* **105**, 11050–11054 (2008).
27. Schindler, P. et al. Experimental repetitive quantum error correction. *Science* **332**, 1059–1061 (2011).
28. Knill, E., Laflamme, R., Martinez, R. & Negrevergne, C. Benchmarking quantum computers: the five-qubit error correcting code. *Phys. Rev. Lett.* **86**, 5811–5814 (2001).

29. DiCarlo, L. *et al.* Preparation and measurement of three-qubit entanglement in a superconducting circuit. *Nature* **467,** 574–578 (2010).
30. Vitali, D., Zippili, S., Tombesi, P. & Raimond, J. M. Decoherence control with fully quantum feedback schemes. *J. Mod. Opt.* **51,** 799–809 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

LETTER

Increased forest ecosystem carbon and nitrogen storage from nitrogen rich bedrock

Scott L. Morford[1], Benjamin Z. Houlton[1] & Randy A. Dahlgren[1]

Nitrogen (N) limits the productivity of many ecosystems worldwide, thereby restricting the ability of terrestrial ecosystems to offset the effects of rising atmospheric CO_2 emissions naturally[1,2]. Understanding input pathways of bioavailable N is therefore paramount for predicting carbon (C) storage on land, particularly in temperate and boreal forests[3,4]. Paradigms of nutrient cycling and limitation posit that new N enters terrestrial ecosystems solely from the atmosphere. Here we show that bedrock comprises a hitherto overlooked source of ecologically available N to forests. We report that the N content of soils and forest foliage on N-rich metasedimentary rocks (350–950 mg N kg^{-1}) is elevated by more than 50% compared with similar temperate forest sites underlain by N-poor igneous parent material (30–70 mg N kg^{-1}). Natural abundance N isotopes attribute this difference to rock-derived N: ^{15}N/^{14}N values for rock, soils and plants are indistinguishable in sites underlain by N-rich lithology, in marked contrast to sites on N-poor substrates. Furthermore, forests associated with N-rich parent material contain on average 42% more carbon in above-ground tree biomass and 60% more carbon in the upper 30 cm of the soil than similar sites underlain by N-poor rocks. Our results raise the possibility that bedrock N input may represent an important and overlooked component of ecosystem N and C cycling elsewhere.

Globally, sedimentary rocks contain 10^{21} g of fixed N, considerably more than the 10^{19} g of fixed N in the total biosphere[5]. Such N is primarily derived from the burial of organic matter in marine and freshwater sediments, where it is incorporated into rock as organic N or as ammonium in silicate minerals. Sedimentary and metasedimentary rocks are distributed globally and typically contain between 200 and $1,200$ mg N kg^{-1}, whereas high-grade metamorphic and igneous rocks typically contain less than 40 mg N kg^{-1} (ref. 6). Fixed N is also found as nitrate salts in arid environments[7], owing to the deposition of atmospheric N over millennia.

Reservoirs of silicate N were identified more than half a century ago[8], but it is generally believed that rock N is not sufficiently important to alter the terrestrial N cycle, despite evidence to the contrary[9]. Just as weathering of phosphorus (P) from bedrock is considered the dominant source of P to terrestrial ecosystems[10], geological N may also be a long-term source of bioavailable N to plants and ecosystems. Bedrock has already been implicated as a source of N to aquifers[11] and surface waters[12]. Here we show, using a range of chemical and isotopic techniques, that bedrock contributes substantial amounts of N to temperate coniferous forests. In addition, we conducted a regional-scale investigation of 88 forest inventory and analysis (FIA) plots to show that total forest carbon storage is higher in ecosystems underlain by N-rich bedrock than in those underlain by N-poor rocks.

We tested the hypothesis that rock weathering is an ecologically significant N source that can increase forest productivity and C storage in temperate coniferous forests of northern California, USA. The first site, South Fork Mountain (SFM), is underlain by mica schist derived from low-grade metamorphism of marine sediments dating to the early Cretaceous period[13]. The mica phase of the schist contains at least $2,700$ mg N kg^{-1} as interlayer ammonium that is released to soil

solution as the rock weathers[14]. Our second site, adjacent to SFM, is underlain by the Bear Wallow Diorite Complex (BWDC), a plutonic rock dating to the Jurassic period[13].

These sites have common tectonic histories, with significant uplift initiated during the Pleistocene[15] and modern uplift rates estimated to be from 1 to 4 mm yr^{-1} (see Methods). Other than parent material, all other state factors are similar across sites (Table 1). Soils at both sites are classified as Dystroxerepts, have sandy loam to loam texture, and show soil development to a depth of 0.5–1.0 m. The strontium isotope (^{87}Sr/^{86}Sr) composition of foliage from SFM and BWDC suggests that the two sites receive similar fractions of Sr inputs from precipitation and bedrock (Supplementary Fig. 1). Although differences in mineralogy and Sr concentration preclude a direct comparison of bedrock weathering inputs between sites[16], the foliar Sr data are consistent with other soil metrics in suggesting that the weathering status of our sites is comparable. In addition, stand composition and tree age at SFM (tree age 121 ± 7.5 years, mean $\pm$ s.e.m.) and BWDC (111 ± 7.6) are similar, and the ^{13}C/^{12}C of foliage is largely indistinguishable between sites (Supplementary Table 1), indicating similar plant-available water regimes. Atmospheric deposition contributes less than 1 kg N ha^{-1} yr^{-1} (ref. 17), whereas biological N_2-fixation inputs are estimated to be about 5 kg ha^{-1} yr^{-1}, on the basis of measurements from similar mature forests in western Oregon[18], empirical models[19] and numerical simulations[3].

In contrast, the mean N content of bedrock differs substantially: 682 ± 50 and 55 ± 6 mg N kg^{-1} (mean $\pm$ s.e.m.) at SFM and BWDC, respectively (Fig. 1a). The bulk-N content of the SFM rock is similar to the global average for marine pelagic sediments, but is lower than for most organic-rich marine sediments, which commonly exceed $1,000$ mg N kg^{-1} (ref. 20). The differences in rock N pools translate to profoundly different N cycles: surface mineral soils in SFM forests

Table 1 | State factors and soil characteristics for forests on SFM and BWDC

Parameter	SFM	BWDC
State factors		
Parent material	Mica schist	Diorite–gabbro
Elevation (m)	1,650–1,720	1,400–1,500
Aspect	N–NE	N–NE
Precipitation (mm)	1,520	1,400
Mean annual temperature (°C)	9	10
Dominant soil type	Dystroxerept	Dystroxerept
Mineral soil characteristics (0–30 cm)		
Soil texture	Loam–sandy loam	Sandy loam
Soil pH (1:1 soil/water paste)	4.85	5.62
Bulk density (g cm^{-3})	1.05	1.09
Coarse fragments (%)	25	30
Clay content (%)	8–12	5–8
Total C (mass %)	5.51 ± 0.30	3.54 ± 0.34
C/N (mol/mol)	20.8 ± 1.0	29.5 ± 1.3
Soil C storage (Mg ha^{-1})	130.2 ± 7.1	81.0 ± 7.8
Soil N storage (Mg ha^{-1})	7.3 ± 0.35	3.2 ± 0.26

Soil physical and chemical characteristics in the top 30 cm of mineral soil are reported. C and N pools in soil are reported as means $\pm$ s.e.m. ($n = 34$).

[1]Department of Land, Air and Water Resources. University of California – Davis, California 95616, USA.
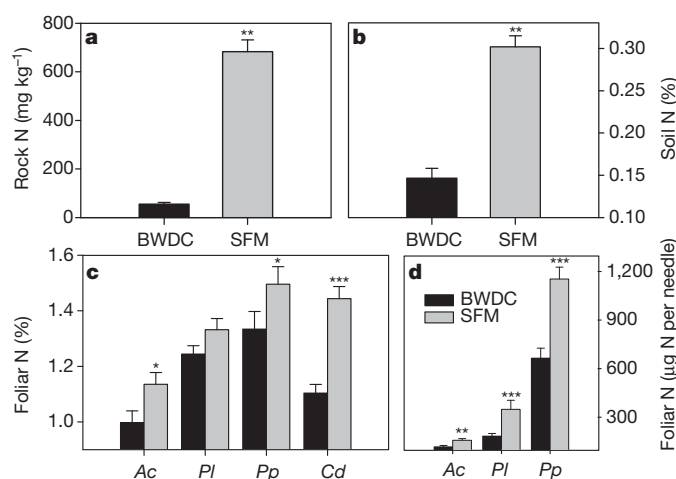
Figure 1 | **Total nitrogen in rock, soil and foliage pools for SFM and BWDC forests. a–c,** Total nitrogen in rock (mg N kg^{-1}, $n = 18$) (**a**), soil (%N, $n = 34$) (**b**) and plant foliage (%N, $n = 80$) (**c**) in the BWDC (black) and SFM (grey) forests. **d,** Foliar nitrogen expressed in µg N per needle, to account for biomass dilution. *Calocedrus decurrens* is not presented in **d** because of scale-leaf rather than needle-leaf morphology. Error bars represent s.e.m. Species sampled: *Ac, Abies concolor; Pl, Pinus lambertiana; Pp, Pinus ponderosa; Cd, Calocedrus decurrens.* Asterisk, $P < 0.05$; two asterisks, $P < 0.01$; three asterisks, $P < 0.001$.

Figure 2 | **Nitrogen isotope values of the rock–soil–plant system. a,** SFM forest; **b,** BWDC forest. The median and range are plotted for each component. The dashed line represents the approximate isotope value for total atmospheric N inputs.

contain significantly more N than those in BWDC forests ($3,026 \pm 149$ and $1,426 \pm 113$ mg N kg^{-1} (mean ± s.e.m.), respectively; Fig. 1b). In addition, soil C/N ratios in the top 30 cm of mineral soil are significantly lower at SFM (20.8 mol:mol) than at BWDC (29.5). Elevated total C concentrations in SFM soil result in substantially more C storage in the top 30 cm of soil at SFM (130 ± 7 Mg C ha^{-1}; mean ± s.e.m.) than in BWDC (81 ± 8 Mg C ha^{-1}; Table 1). These differences are consistent with a California statewide soils database ($n = 183$; Supplementary Fig. 2); C and N storage at BWDC is similar to the state average for comparable ecosystems (74.9 Mg C ha^{-1}, 3.89 Mg N ha^{-1}), whereas SFM has among the highest C and N contents observed.

Further, N enrichment in rocks and soils is readily apparent in the tree foliage. On average, conifer needles at SFM contain 50% more N per needle than at BWDC; this difference is observed in three of four sampled species (Fig. 1c); trees on SFM contain 7–30% more N on a foliar mass basis. *Calocedrus decurrens* was the most N-enriched: $1.45 \pm 0.04\%$ N (mean ± s.e.m.) at SFM, versus $1.10 \pm 0.03\%$ N at BWDC. The other trees showed similar but less substantial N enrichment (*Abies concolor*, $1.14 \pm 0.04\%$ N and $1.00 \pm 0.04\%$ N; *Pinus ponderosa*, $1.50 \pm 0.06\%$ N and $1.33 \pm 0.06\%$ N; *Pinus lambertiana*, $1.33 \pm 0.04\%$ N and $1.25 \pm 0.03\%$ N at SFM and BWDC, respectively).

Under N-rich conditions, plants can either concentrate N in foliage (for example, *Calocedrus*) or build more foliage (or do both; for example *Abies* and *Pinus*), the former leading to higher N concentrations in foliage, the latter to no major changes in N contents. At SFM, needle biomass in pines is roughly 70% higher than at BWDC (Supplementary Table 2), implying significant gains in biomass with added N. Examining N contents on a per-needle basis to account for nutrient dilution[21] reveals even more profound differences between sites: SFM trees show 35–90% more N than those at BWDC (Fig. 1d). It therefore seems that forest responses to geological N inputs are manifested as higher foliar biomass production, in addition to the higher foliar N concentrations observed.

Next, we used natural-abundance ^{15}N/^{14}N stable isotopes (δ^{15}N parts per thousand (‰) versus air) to trace the movement of N from rocks to soils and plants (Fig. 2). Nitrogen-isotope data for the BWDC site are consistent with expectations for other similar N-limited temperate conifer forests[22,23]: foliar δ^{15}N (-1.59 ± 0.19‰; mean ± s.e.m.) is depleted relative to soil (1.96 ± 0.32‰), whereas δ^{15}N in rocks (18.37 ± 1.55‰) is substantially elevated compared with that in plants
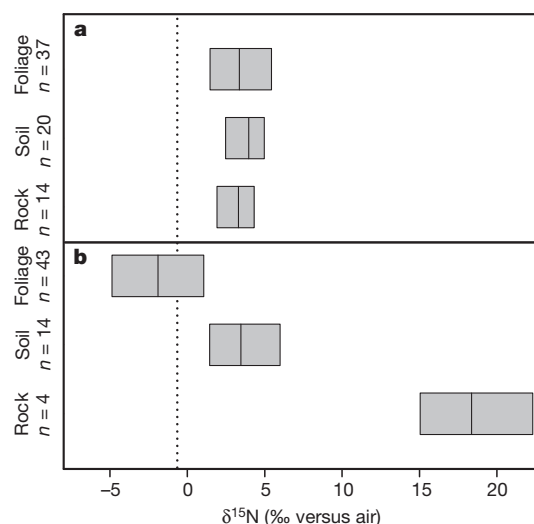
and soils. In contrast, foliar δ^{15}N at SFM (3.10 ± 0.17‰) is 4.7‰ higher than at BWDC and is essentially indistinguishable from that of soils (3.6 ± 0.17‰) and rock (3.3 ± 0.22‰). This points to a direct link between weathering of rock N and the elevated N status of SFM, and conforms with general patterns of plant and soil δ^{15}N in N-rich ecosystems[23]. Moreover, the lack of differentiation between the ^{15}N/^{14}N of rock inputs and soils at SFM indicates little or no isotopic expression through N losses at SFM. This agrees with observations of high-nitrate leachates in soils[14] and streams[12] associated with N-rich rock substrates, given that N leaching does not seem to impart a major fractionation of N isotopes in comparison with gaseous losses of N (ref. 24). Finally, the contrast of plant and soil δ^{15}N pools at SFM and BWDC is consistent with models that show increased isotopic expression of ectomycorrhizal N transfer under low availability of N (ref. 25) (that is, at BWDC), although we do not observe differences in foliar δ^{15}N between ectomycorrhizal and arbuscular mycorrhizal species at either site (Supplementary Fig. 3).

Rising levels of atmospheric CO_2 and climate change have renewed interests in links between N and C cycles, especially in high-latitude forests[4]. Nitrogen limitation of extra-tropical forests is widespread[26]; new N inputs therefore have the potential to allow for more CO_2 uptake and storage on land[27], thus affecting the pace and magnitude of global climate change. Using FIA data from forests developing on bedrock similar to that at SFM (eastern Franciscan, N-rich) and BWDC (western Klamath, N-poor) geological provinces of Northwest California, we show that geological N may also contribute to enhanced forest productivity and C storage on land.

Holding other state factors constant, our analysis suggests that forests on bedrock with high N-enrichment potentials (metapelites formed under a low geothermal gradient), on and within the SFM region, contain 42% more C in above-ground tree biomass than similar forests in the region that are underlain by N-depleted bedrock ($P = 0.009$, $R^2 = 0.66$; Fig. 3). Although lithology in our model represents the sum of the parent material influences on tree C stocks, the physical properties inherited from parent material, such as soil texture and depth, are taken into account by the available water-holding capacity parameter ($P = 0.001$). This implies that the enhanced C storage on putatively N-rich lithology is most associated with differences in nutrient release from rocks. Given the N-limited propensity of such forests, these region-wide results indicate higher productivity and C storage owing to bedrock N, which is consistent with the higher N and C stocks at SFM (Table 1 and Fig. 1) and with Forest Service data suggesting that
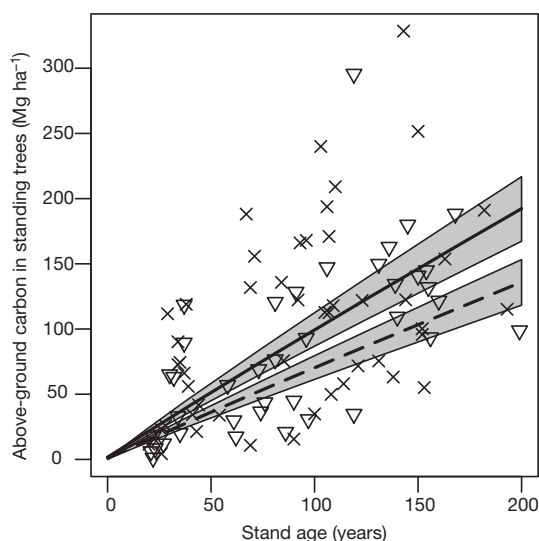
Figure 3 | Carbon in above-ground tree biomass for forests growing on N-rich and N-poor lithology. Plot of carbon ($Mg\,ha^{-1}$) against stand age for FIA plots ($n = 88$) on N-rich (crosses) and N-poor (triangles) lithologies used in the model. To account for model axes not shown, we present detransformed model estimates for sites on N-rich (solid line) and N-poor (dashed line) lithologies, holding other model parameters constant at nominal values to illustrate differences in carbon storage attributable to lithology. The grey area indicates lithology s.e.m. Our analysis estimates that sites on N-rich lithology contain 42% more C in above-ground tree biomass than sites on N-poor lithology ($P = 0.009$; adjusted $R^2 = 0.66$) after accounting for confounding state factors.

SFM contains the most productive Douglas fir forests remaining in California[28].

Our results raise the possibility that rock weathering may be a significant source of N to terrestrial ecosystems underlain by N-rich substrates elsewhere. There are many documented cases in which the N budgets of forests are out of balance, often being in need of substantially higher N input rates to account for N accumulation in soil and plants[29]. In cases where bedrock is enriched in N, parent-material N could explain some (or all) of the missing N inputs. For example, using the N isotope data to devise a set of mass balance models, we estimate that rock N sources contribute 30–100% of the ecosystem N inputs at SFM, depending on the degree of isotopic expression by N losses (Supplementary Table 3). These results are comparable to simple uplift models (rock $N = 47\%$ of N inputs) and weathering experiments in the laboratory (rock $N = 64\%$ of N inputs). In terms of fluxes, these various approaches point to substantial N inputs: 3.0–$10.9\,kg\,ha^{-1}\,yr^{-1}$ by rock weathering at SFM, potentially more than doubling known inputs from the atmosphere (that is, fixation plus deposition equal to $6\,kg\,N\,ha^{-1}\,yr^{-1}$ (refs 3, 17–19)). Given that sedimentary and metasedimentary rocks contain 99% of the global fixed N and cover roughly 75% of the Earth's land surface[30], the potential for bedrock N to stimulate productivity and C storage in the terrestrial biosphere seems globally significant.

METHODS SUMMARY

One-year-old sun foliage from the mid-canopy of mature conifer trees was collected with a pole saw in autumn 2008, after full needle elongation. Foliage was rinsed with deionized water and dried at $50\,^{\circ}C$ for 48 h. Needle biomass was determined by a 100-needle count on dried foliage. Samples were ground and analysed for total C and N, $\delta^{15}N$ and $\delta^{13}C$ by the University of California – Davis Stable Isotope Facility with a SerCon Hydra 20/20 isotope ratio mass spectrometer (IRMS). Isotope values are reported in parts per thousand (‰) relative to Vienna PeeDee Belemnite for $^{13}C/^{12}C$, and air N_2 for $^{15}N/^{14}N$, using standard delta (δ) notation. Surface mineral soils were collected to a depth of 30 cm, dried in air, sieved to 2 mm, milled to pass a 200-µm sieve, and then analysed by IRMS for total C and N, $\delta^{15}N$ and $\delta^{13}C$. Bulk density was determined by the core method, and soil texture by laser diffraction. Minimally weathered bedrock samples were collected

from outcrops at each site, cut with a slab saw to remove weathered surfaces, treated with 5% hydrogen peroxide for 24 h to remove any remnant surficial organic matter, and pulverized with a carbide-steel shatter box to pass a 75-µm sieve. Samples were analysed by IRMS for total C and N, $\delta^{15}N$ and $\delta^{13}C$. FIA plot data, including intensification plots, were obtained from the United States Forest Service. Above-ground carbon stocks were calculated with the Jenkins equations in the 'Fire and fuels extension' of the Forest Vegetation Simulator. Student's t-tests were used to compare differences between sites in total N in rocks, soil and foliage. The FIA model took the form of a multiple linear regression. Statistical tests and model fitting were performed in R.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Vitousek, P. M. & Howarth, R. W. Nitrogen limitation on land and in the sea: how can it occur? *Biogeochemistry* **13**, 87–115 (1991).
2. Oren, R. *et al.* Soil fertility limits carbon sequestration by forest ecosystems in a CO_2-enriched atmosphere. *Nature* **411**, 469–472 (2001).
3. Wang, Y. P. & Houlton, B. Z. Nitrogen constraints on terrestrial carbon uptake: implications for the global carbon–climate feedback. *Geophys. Res. Lett.* **36**, L24403, doi:10.1029/2009gl041009 (2009).
4. Magnani, F. *et al.* The human footprint in the carbon cycle of temperate and boreal forests. *Nature* **447**, 848–850 (2007).
5. Galloway, J. N. in *Treatise on Geochemistry* Vol. 8 (ed. W. H. Schlesinger) 557–583 (Elsevier Pergamon, 2004).
6. Holloway, J. M. & Dahlgren, R. A. Nitrogen in rock: occurrences and biogeochemical implications. *Glob. Biogeochem. Cycles* **16**, 1118, doi:10.1029/2002gb001862 (2002).
7. Walvoord, M. A. *et al.* A reservoir of nitrate beneath desert soils. *Science* **302**, 1021–1024 (2003).
8. Stevenson, F. J. On the presence of fixed ammonium in rocks. *Science* **130**, 221–222 (1959).
9. Cornwell, S. M. & Stone, E. L. Availability of nitrogen to plants in acid coal mine spoils. *Nature* **217**, 768–769 (1968).
10. Walker, T. W. & Syers, J. K. The fate of phosphorus during pedogenesis. *Geoderma* **15**, 1–19 (1976).
11. Strathouse, S. M., Sposito, G., Sullivan, P. J. & Lund, L. J. Geologic nitrogen—a potential geochemical hazard in the San Joaquin Valley, California. *J. Environ. Qual.* **9**, 54–60 (1980).
12. Holloway, J. M., Dahlgren, R. A., Hansen, B. & Casey, W. H. Contribution of bedrock nitrogen to high nitrate concentrations in stream water. *Nature* **395**, 785–788 (1998).
13. Irwin, W., Wolfe, E. W., Blake, M. C. & Cunningham, C. G. *Geologic map of the Pickett Peak quadrangle, Trinity County, California* (US Geological Survey Map GQ-1111, 1974).
14. Dahlgren, R. A. Soil acidification and nitrogen saturation from weathering of ammonium-bearing rock. *Nature* **368**, 838–841 (1994).
15. Aalto, K. R. in *Geological Studies in the Klamath Mountains Province, California and Oregon. A Volume in Honor of William P. Irwin* (eds Snoke, A. W. & Barnes, C. G.) 451–464 (Geological Society of America Special Papers, vol. 410, 2006).
16. Capo, R. C., Stewart, B. W. & Chadwick, O. A. Strontium isotopes as tracers of ecosystem processes: theory and methods. *Geoderma* **82**, 197–225 (1998).
17. National Atmospheric Deposition Program. *National Atmospheric Deposition Program 2008 Annual Summary* (NADP Data Report 2009–01) (Univ. of Illinois at Urbana-Champaign, 2009).
18. Sollins, P., Grier, C. C., McCorison, F. M., Cromack, K. & Fogel, R. The internal element cycles of an old-growth Douglas-fir ecosystem in western Oregon. *Ecol. Monogr.* **50**, 261–285 (1980).
19. Cleveland, C. C. *et al.* Global patterns of terrestrial biological nitrogen (N_2) fixation in natural ecosystems. *Glob. Biogeochem. Cycles* **13**, 623–645 (1999).
20. Li, Y. H. Distribution patterns of the elements in the ocean—a synthesis. *Geochim. Cosmochim. Acta* **55**, 3223–3240 (1991).
21. Timmer, V. R. & Stone, E. L. Comparative foliar analysis of young Balsam fir fertilized with nitrogen, phosphorus, potassium, and lime. *Soil Sci. Soc. Am. J.* **42**, 125–130 (1978).
22. Amundson, R. *et al.* Global patterns of the isotopic composition of soil and plant nitrogen. *Glob. Biogeochem. Cycles* **17**, 10.1029/2002gb001903 (2003).
23. Martinelli, L. A. *et al.* Nitrogen stable isotopic composition of leaves and soil: tropical versus temperate forests. *Biogeochemistry* **46**, 45–65 (1999).
24. Houlton, B. Z. & Bai, E. Imprint of denitrifying bacteria on the global terrestrial biosphere. *Proc. Natl Acad. Sci. USA* **106**, 21713–21716 (2009).
25. Hobbie, E. A., Macko, S. A. & Williams, M. Correlations between foliar delta N-15 and nitrogen concentrations may indicate plant–mycorrhizal interactions. *Oecologia* **122**, 273–283 (2000).
26. LeBauer, D. S. & Treseder, K. K. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology* **89**, 371–379 (2008).
27. Thomas, R. Q., Canham, C. D., Weathers, K. C. & Goodale, C. L. Increased tree carbon storage in response to nitrogen deposition in the US. *Nature Geosci.* **3**, 13–17 (2010).
28. Keeler-Wolf, T. *Ecological surveys of Forest Service Research Natural Areas in California* (Gen. Tech. Rep. PSW-125) (Pacific Southwest Research Station, Forest Service, 1990).

29. Galloway, J. N., Schlesinger, W. H., Levy, H. II, Michaels, A. & Schnoor, J. L. Nitrogen fixation: anthropogenic enhancement–environmental response. *Glob. Biogeochem. Cycles* **9,** 235–252 (1995).
30. Blatt, H. & Jones, R. L. Proportions of exposed igneous, metamorphic, and sedimentary rocks. *Geol. Soc. Am. Bull.* **86,** 1085–1088 (1975).

Author Contributions S.L.M., B.Z.H. and R.A.D. contributed to the experimental design and collection of field samples. S.L.M. performed sample processing and laboratory analysis. S.L.M. designed and implemented the FIA modelling component. S.L.M., B.Z.H. and R.A.D. contributed to the interpretation of laboratory data, modelling results and manuscript preparation.

METHODS

Field sampling and laboratory analysis. Sampling areas were located near 40° 18′ N 123° 18′ W and 40° 26′ N 123° 21′ W for the SFM and BWDC sites, respectively. Within each sampling area, three locations were chosen at random for collection of foliage, soils and rock. A pole saw was used to collect sun needles from 10 m above the forest floor from co-dominant conifer trees in the autumn of 2008. Three clippings from each tree were bulked into a single sample. One-year-old needles were removed from the branches by hand, and needles showing damage, discoloration or herbivory were discarded. The separated needles were washed with deionized water and then dried for 48 h at 50 °C. A subsample of dried foliage was ground with a Wiley mill and then pulverized with a ball mill before chemical analysis. From the remaining sample, 100 needles were chosen at random and weighed to determine a 100-needle mass.

Ground samples were placed into tin capsules and analysed for total C and N, $^{13}C/^{12}C$ and $^{15}N/^{14}N$ on a Sercon 20/20 elemental analysis–IRMS at the University of California – Davis Stable Isotope Facility. Isotope values are reported in parts per thousand (‰) relative to Vienna PeeDee Belemnite for $^{13}C/^{12}C$ and air N_2 for $^{15}N/^{14}N$, using standard delta (δ) notation[31]. Foliage replicates had standard deviations for total C, total N, $\delta^{13}C$ and $\delta^{15}N$ of less than 0.1%, 0.05%, 0.05‰ and 0.2‰, respectively.

We dug three 50 cm × 50 cm pits in each sampling location down to the C_r horizon (60–80 cm) to describe soil taxonomy. Additional smaller pits (0.3 m × 0.3 m) were used to collect representative mineral soil samples for the top 30 cm of the profile for chemical analysis. Soil bulk density and volume of coarse fragments were quantified in the surface mineral horizons with the use of the core method[32]. Soil texture was determined by laser diffraction[33]. Soils were air-dried in the laboratory and sieved to 2 mm; fine roots were then removed with forceps. The soil fine fraction was ground with a ball mill, weighed into tin capsules and analysed for elemental and isotopic analysis (see foliar methods for details). Soil replicates had standard deviations for total C, total N and $\delta^{15}N$ of less than 0.02%, 0.005% and 0.3‰, respectively.

Rock samples were recovered from minimally weathered outcrops with a crack hammer. The weathering rind was removed with a lapidary slab saw. The sample was then treated with 5% hydrogen peroxide solution for 24 h to remove surficial organic matter, dried and then crushed with a hydraulic press to obtain particles with diameters of less than 10 mm. Samples were then washed with deionized water and dried at 110 °C for 48 h. A 60 g subsample was pulverized with a carbide-steel shatter box to pass a standard US 200 mesh (74-μm) sieve. Samples were weighed into tin capsules and submitted to the Stable Isotope Facility for elemental and isotopic analysis. Replicates for SFM schist had standard deviations for total N and $\delta^{15}N$ of 10 mg kg^{-1} and 0.3‰, respectively. Replicates for rock samples from the BWDC site had higher analytical uncertainty, with standard deviations of total N of 15 mg kg^{-1} and a $\delta^{15}N$ of 2‰. In addition, two independent methods were used to quantify total N in a subset of samples: vanadium oxide catalysed combustion[34] and HF/HCl digestion followed by conductimetric NH_4 quantification[35]. Both methods resulted in N recovery within 10% of the standard elemental analysis–IRMS method.

Comparison of sample means of total N in rock, soil, and foliage between SFM and BWDC sites was performed with Student's t-test in R[36]. Homogeneity of variance was tested with the Sharpio–Wilk test for normality.

For Sr analysis, rock and foliage samples were prepared as described above; 0.25 g of foliage was digested with ultrapure nitric acid and hydrogen peroxide in a CEM MDS-2100 microwave digester. Snow was collected immediately after a storm in March 2009 and passed through pre-rinsed 0.45-μm Millipore Millix syringe filters. Sr extractions for all samples were performed with Eicrhom Sr-Spec resin and analysed on the Nu Plasma HR multi collector–inductively coupled plasma–mass spectrometer (MC–ICP–MS) at the University of California – Davis Interdisciplinary Center for Plasma Mass Spectrometry. Sr was corrected to $^{86}Sr/^{88}Sr = 0.1194$ to account for instrumental mass fractionation, and then normalized to SRM 987 ($^{87}Sr/^{86}Sr = 0.710248$). The SRM 987 standard averaged $^{87}Sr/^{86}Sr = 0.710269 \pm 0.000040$ (2 s.d., $n = 25$). Method blanks were less than 0.4 and 1.3 ng for rock and foliage, respectively.

Analysis of FIA plot data. A total of 130 FIA plots were identified within the designated study area and included both the standard and intensification plots for rare forest types[37–39]. Potential N-bearing lithologies were identified with available geological data[13,40–45], field observations and geochemistry data from our rock collection. Potential N-rich sites occur on metasediments (blueschist facies) in the eastern Franciscan belt, whereas N-poor sites are located in terranes of western Klamath that have undergone substantial contact metamorphism. Of the 130 sites, 89 were initially selected for inclusion in the analysis by using the following criteria:

1. We selected for mixed pine, Douglas fir, Douglas fir–white fir, and Douglas fir–ponderosa pine forest types by using CalVeg classification data[46].

2. Plots occurring on ultramafic/serpentine soils were excluded by using spatial data from the Trinity serpentine soil survey[47].

3. Plots were located on slopes less than 80% and had mean annual precipitation values from 1,000 to 2,000 mm.

4. Plots had a mean stand age (defined below) of more than 20 years and less than 200 years.

5. The Stand Visualization System (SVS)[48] was used to prescreen for stands that exhibited significant mortality in the largest diameter classes resulting from extreme disturbance events such as high-intensity wildfires or insect and microbial pathogens. Stands in which more than 75% of the canopy area was occupied by dead standing trees in combination with 50–80% mortality in the largest tree diameter classes were removed from the analysis.

A total of 88 plots were included in the final analysis (Supplementary Table 4); one was removed after an analysis of disturbance regimes using fire history and stand density characteristics (discussed below).

Seven parameters were used as potential explanatory variables for the analysis. Supplementary Table 5 provides information on the central tendency and data distribution for each parameter. Non-significant variables ($P > 0.05$) were dropped from the final analysis during model fitting (see discussion below). The full set of independent parameters tested in the analysis includes:

1. Stand age. The FIA intensification plot database does not include information on stand age; we therefore estimated stand age (A_s) using a basal area-weighted function of measured conifer trees within each plot, calculated from

$$A_s = \sum A_t A_b / \sum A_b \qquad (1)$$

where A_t is the age of an individual measured tree, and A_b is a calculated value representing the basal area per hectare. Age data from live, measured conifer trees only were only used in this determination.

2. Slope. Percentage slope was taken directly from the FIA database and converted to decimal fractions. In the case of a split plot, the mean slope across the four subplots was used.

3. Available water-holding capacity (AWHC). An estimate of the total amount of water the soil profile can hold, based on soil texture, depth and coarse fragments (fraction larger than 2 mm in diameter). AWHC values are spatially averaged values of SSURGO and STATSGO data, using a 1-km^2 grid[49].

4. Insolation. Annual solar input (W h m^{-2}) was derived from the 1/3-arcsecond (10-m) National Elevation Data set (NED)[50] using the Solar Analyst function[51] in ArcGIS 9.3 (ref. 52). Insolation values were linearly rescaled from 0 to 1.

5. Precipitation. Mean annual precipitation was derived from the PRISM 30-arcsecond (800-m) data set (1971–2000)[53].

6. Temperature. Mean annual temperature was derived from the PRISM 30-arcsecond (800-m) data set (1971–2000)[54].

7. Lithology. Categorical variable that indicates whether the plot is located on lithologies with nitrogen statuses similar to SFM (eastern Franciscan) or BWDC (western Klamath).

The response variable in the model was above-ground carbon (C_a) in standing live trees (Mg C ha^{-1}). C_a was calculated on a per-tree basis by using Jenkins allometric biomass equations[55] in the 'Fire and fuels' extension[56] of the Forest Vegetation Simulator (FVS)[57].

To account for potential differences in disturbance regimes between regions, we used fire history[58] in combination with stand density information to identify plots where disturbance had substantially influenced carbon storage. A total of 26 of the initial 89 plots had a record of fire since 1920 (29%), with 11 recorded incidents in the eastern Franciscan and 15 in the western Klamath. Only 14 of the 26 fires occurred within the past 50 years, with 40% occurring in the eastern Franciscan. Inclusion of a categorical parameter in the model to account for the presence or absence of fire was non-significant ($P > 0.05$) at the three timescales examined (15, 30 and 50 years).

We compared stand density with basal area across plots to identify sites where fire significantly altered stand structure. To compare sites directly, we calculated Reineke's stand density index (SDI)[59] for each plot to account for the negative exponential relationship (reverse J-shaped distribution) between tree size and density[60]. After a log–log transformation to account for data heteroscedasticity, we found a strong linear relationship between log(SDI) and logC_a across lithologies and fire histories ($R^2 = 0.92$; Supplementary Fig. 4). We identified and removed a single plot that deviated significantly from the population SDI:A_b trend. The removed plot, located in the western Klamath (N-poor lithology), was 194 years old with only 59 Mg C ha^{-1} in above-ground tree biomass.

We used R[36] to fit a multiple linear regression model, accounting for two-way interactions between local state factors. Initial model parameter selection was performed with the 'step' function in R—an implementation of Akaike's information criterion[61]. In favour of parsimony, the final model included only parameters

deemed significant ($P \leq 0.05$) by type II analysis of variance with the 'Anova' function from the Companion to Applied Regression (CAR)[62] library in R[36]. The final model took the form

$$\log C_a = \beta_0 + \log x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + \varepsilon \quad (2)$$

where β_0 is the intercept, x_1–x_6 represent coefficients fitted by the model (x_1, stand age; x_2, lithology, x_3, temperature, x_4, AWHC; x_5, insolation; x_6, AWHC × insolation) and ε represents error. An ANOVA table for the final model is presented in Supplementary Table 6). Precipitation ($P = 0.12$) and slope ($P = 0.49$) were found to be non-significant in the final model. Collinearity was assessed by using the variance inflation factor from the CAR package[62] (Supplementary Table 7). Model coefficient tables (Supplementary Table 7), model fit information (Supplementary Fig. 5) and residual plots (Supplementary Fig. 6) are also presented.

Mass balance modelling methods for determination of bedrock N flux. We developed four simple models derived from natural-abundance N isotope data, tectonic uplift rates and laboratory weathering experiments to better constrain N input fluxes from bedrock into our sites. Our first two models apply principles of isotopic mass balance: model A estimates rock N contribution to the ecosystem under the assumption of no isotopic fractionating loss, whereas model B assumes that fractionating loss pathways are similar among sites. Model C uses steady-state assumptions to estimate N inputs using rock N content, tectonic uplift rates, and chemical weathering potential. Finally, model D uses laboratory weathering experiments, corrected for temperature with Arrhenius relationships. All models were implemented in R and results are based on 100,000 Monte Carlo simulations. We report 2.5th centiles, medians and 97.5th centiles of all model outputs, representing a 95% confidence interval around our estimates.

In model A we employ a simple mixing model to estimate the fraction of rock N inputs to the ecosystems (f_{rock}, equation (3)), using measured isotopic data and estimated pool sizes. The major assumption in this model is that there is no fractionating loss at either the BWDC or SFM site. We estimate an ecosystem $^{15}N/^{14}N$ ($\delta^{15}N_{ecosystem}$) pool through the mixing of soil (f_{soil}) and biomass (f_{plant}) pools. The relative size of the soil versus biomass pools is difficult to quantify for these sites; however, we can use estimates of forest ecosystem carbon pools[63] and the C/N of plant biomass pools[3,64,65] to predict a range of possible relative pool sizes. Using these constraints, we calculate that 0.07–0.40 of the ecosystem N reservoir is contained within plant biomass, but larger values most probably represent forests where soil C and N storage is low. In our model we use a range of 0.07–0.20 for the biomass contribution to the total ecosystem N pool, and assume that the $^{15}N/^{14}N$ of foliage is representative of the entire biomass pool ($\delta^{15}N_{plants}$, equation (4)). The $^{15}N/^{14}N$ of the atmospheric endmember ($\delta^{15}N_{atm}$) is assumed to range from -1.5‰ to -0.5‰, and $^{15}N/^{14}N$ of the rock ($\delta^{15}N_{rock}$) and soil ($\delta^{15}N_{soil}$) components in the model represent the mean ± s.e.m. for each pool.

$$f_{rock} = (\delta^{15}N_{ecosystem} - \delta^{15}N_{atm})/(\delta^{15}N_{rock} - \delta^{15}N_{atm}) \quad (3)$$

$$\delta^{15}N_{ecosystem} = f_{soil}(\delta^{15}N_{soil}) + f_{plants}(\delta^{15}N_{plants}) \quad (4)$$

Model B differs from model A in that it incorporates a fractionating loss term whereby isotopically light N is removed preferentially from the ecosystem (equation (5)). Loss pathways include both leaching ($f_{leaching}$) and denitrification (f_{denit}); however, evidence indicates that the isotope effect of leaching is approximately zero when integrated through space and time[24] (equation (6)), leaving denitrification as the sole fractionating loss pathway in the ecosystem. To consider isotopic effects of N losses, we make two simplifying assumptions. First, we assume that the contribution of N from rock is effectively zero at the BWDC forest. Second, we assume that the relative imprint of denitrification on the total ecosystem pool is the same between sites, given the similarity of climate, vegetation and soil physical properties between SFM and BWDC. In doing so, we combine the fraction and isotope term in the denitrification pathway into a single variable, η (equation (7)). Using the assumptions from equations (6) and (7), we can solve for η (equation (8)) and the fraction of rock inputs at the SFM site (equation (9)):

$$\delta^{15}N_{ecosystem} = f_{atm}(\delta^{15}N_{atm}) + f_{rock}(\delta^{15}N_{rock}) - f_{denit}(\varepsilon_{denit}) - f_{leaching}(\varepsilon_{leaching}) \quad (5)$$

$$\varepsilon_{leaching} = 0 \quad (6)$$

$$\eta = f_{denit}(\varepsilon_{denit}) \quad (7)$$

$$\eta = \delta^{15}N_{atm} - \delta^{15}N_{BWDC} \quad (8)$$

$$\frac{\delta^{15}N_{SFM} + \eta - \delta^{15}N_{atm}}{\delta^{15}N_{SFMrock} - \delta^{15}N_{atm}} = \frac{\delta^{15}N_{SFM} - \delta^{15}N_{BWDC}}{\delta^{15}N_{SFMrock} - \delta^{15}N_{atm}} = f_{rock} \quad (9)$$

For model C we estimated the rock N contribution to SFM and BWDC ecosystems from regional uplift rates and atmospheric N inputs. We assumed that uplift equals denudation rates at our sites and that the chemical weathering fraction is between 0.05 and 0.2 of the total denudation rates, on the basis of estimates from the Sierra Nevada[66]. Total atmospheric N inputs are between 4 and 8 kg ha^{-1} yr^{-1} at both sites, on the basis of regional estimates of N fixation[3,17,18] and N deposition[16]. Quaternary uplift rates in the northwestern California are estimated to be between 0.001 and 0.004 m yr^{-1}, on the basis of geodynamic models[67,68] and field data[69,70]. The fraction of rock N input to the ecosystem (f_{rock}) can then be estimated from equation (10):

$$f_{rock} = r_u f_{cw}(10{,}000\,m^2\,ha^{-1})\rho_d\,N_{rock}[1/(1 + i_{atmos})] \quad (10)$$

where r_u is uplift rate (m yr^{-1}), f_{cw} is the chemical weathering fraction, ρ_d is rock density (kg m^{-3}), N_{rock} is the N content of bedrock, and i_{atmos} is the total atmospheric N input (kg ha^{-1} yr^{-1}).

In model D we estimate rock N inputs at SFM on the basis of laboratory weathering data[71]. Bulk leaching experiments estimated bedrock N fluxes to be between 5 and 38 kg^{-1} ha^{-1} yr^{-1} after 360-day incubations at 20 °C. Mineral dissolution rates are sensitive to temperature[30,72,73]; we therefore estimate field N fluxes on the basis of laboratory weathering rates corrected for differences in temperature between the laboratory (20 °C) and field (9 °C), using Arrhenius relationships. Activity coefficients for potassium are used in place of ammonium, given the similarity in charge and ionic radius of ammonium and large-ion lithophiles in geochemical systems[74]. The field weathering rate (r_{field}) is estimated from equation (11):

$$r_{field} = \exp\left[\frac{-E_a}{R}\left(\frac{1}{T_{field}} - \frac{1}{T_{lab}}\right)\right]r_{lab} \quad (11)$$

where E_a is the dissolution activity energy, R is the gas constant, r_{lab} is the laboratory weathering rate, and T_{field} and T_{lab} are field and laboratory temperatures, respectively.

Methods for statewide soil C and N analysis. A total of 183 pedons from the University of California – Davis Soil library were selected for the analysis, on the basis of vegetation data (mixed conifer and Douglas fir). Soils with andic properties were identified and excluded from the analysis by using the database soil taxonomy and through comparison with SSURGO[75] soil survey data at the pedon location. The aqp[76] package in R[36] was used to calculate the depth-integrated C and N content for each pedon at intervals of 1 cm. Calculation of carbon and nitrogen soil storage were based on the horizon-integrated N and C content, the volume of coarse fragments, and the soil bulk density. The bulk density data from the database were incorrect or missing for several pedons; to correct for this, we assumed that bulk density increased linearly from 1,000 kg m^{-3} at the soil surface to 1,200 kg m^{-3} at 30 cm.

31. McKinney, C. R., McCrea, J. M., Epstein, S., Allen, H. A. & Urey, H. C. Improvements in mass spectrometers for the measurement of small differences in isotope abundance ratios. *Rev. Sci. Instrum.* **21,** 724–730 (1950).
32. Blake, G. R. & Hartge, K. H. in *Methods of Soil Analysis—Physical and Mineralogical Methods* Vol. 9 (ed. Klute, A.) Ch. 13 364–367 (Soil Science Society of America, 1986).
33. Eshel, G., Levy, G. J., Mingelgrin, U. & Singer, M. J. Critical evaluation of the use of laser diffraction for particle-size distribution analysis. *Soil Sci. Soc. Am. J.* **68,** 736–743 (2004).
34. Brauer, K. & Hahne, K. Methodical aspects of the N-15-analysis of Precambrian and Palaeozoic sediments rich in organic matter. *Chem. Geol.* **218,** 361–368 (2005).
35. Carlson, R. M. Automated seperaration and conductimetric determination of ammonia and dissolved carbon dioxide. *Anal. Chem.* **50,** 1528–1531 (1978).
36. R Development Core Team. *R: A Language and Environment for Statistical Computing v.2.12.1* (R Foundation for Statistical Computing, 2010).
37. USDA Forest Service. *2007 Mendocino National Forest Vegetation Inventory* (Pacific Southwest Research Station, 2010).
38. USDA Forest Service. *2007 Shasta–Trinity National Forest Vegetation Inventory* (Pacific Southwest Research Station, 2010).
39. USDA Forest Service. *2007 Six Rivers National Forest Vegetation Inventory* (Pacific Southwest Research Station, 2010).
40. Bishop, D. G. South-Fork Mountain Schist at Black-Butte and Cottonwood Creek, northern California. *Geology* **5,** 595–599 (1977).
41. Jayko, A. S. & Blake, M. C. Deformation of the eastern Franciscan belt, northern California. *J. Struct. Geol.* **11,** 375–390 (1989).
42. Lanphere, M. A., Blake, M. C. & Irwin, W. P. Early Cretaceous metamorphic age of South Fork Mountain Schist in northern Coast Ranges of California. *Am. J. Sci.* **278,** 798–815 (1978).

43. Jennings, C. W. & Strand, R. *Geologic Map of California: Ukiah Sheet* (California Division of Mines and Geology, 1960).
44. Strand, R. G. *Geologic Map of California: Redding Sheet* (California Division of Mines and Geology, 1962).
45. Blake, M. S., Helley, E. J., Jayko, A. S., Jones, D. L. & Ohlin, H. N. *Geologic Map of the Willows 1:100,000 Quadrangle, California* (US Geological Survey Open-File Report OF-92-271, 1992).
46. USDA Forest Service. *Existing Vegetation—CALVEG Geodatabase* (USDA Forest Service, Pacific Southwest Region Remote Sensing Lab, 2009).
47. USDA Forest Service. *Trinity Serpentine Soil Survey* (Pacific Southwest Research Station, 2004).
48. McGaughey, R. J. *Stand Visualization System v. 3.3* (United States Department of Agriculture, Forest Service, Pacific Northwest Research Station, 2004).
49. Beaudette, D. E. & O'Geen, A. T. *A 1 km Scale, Gridded Soils Database for California* (California Soil Resource Lab, Univ. of California, Davis, 2010).
50. Gesch, D. B. in *Digital Elevation Model Technologies and Applications: The DEM User Manual* 2nd edn (ed. Maune, D.) 99–118 (American Society for Photogrammetry and Remote Sensing, 2007).
51. Fu, P. & Rich, P. M. A geometric solar radiation model with applications in agriculture and forestry. *Comput. Electron. Agric.* **37,** 25–35 (2002).
52. Environmental Systems Research Group. *ArcGIS Desktop v. 9.3* (Environmental Systems Research Group, Redlands, 2010).
53. The PRISM Group at Oregon State University. *United States Average Monthly or Annual Precipitation, 1971–2000* (Oregon State Univ., 2006).
54. The PRISM Group at Oregon State University. *United States Average Monthly or Annual Maximum Temperature, 1971–2000* (Oregon State Univ., 2006).
55. Jenkins, J. C., Chojnacky, D. C., Heath, L. S. & Birdsey, R. A. National-scale biomass estimators for United States tree species. *For. Sci.* **49,** 12–35 (2003).
56. Reinhardt, E. D., Crookston, N. L. & Rebain, S. A. *The Fire and Fuels Extension to the Forest Vegetation Simulator* (United States Department of Agriculture, Forest Service, 2003).
57. Forest Inventory and Analysis Program. *The Forest Inventory and Analysis Database: Database Description and Users Manual Version 3.0 for Phase 2*, Revision 1. (United States Department of Agriculture, Forest Service, 2008).
58. California Department of Forestry and Fire Protection. *Fire Perimeters (1923 to 2008) for Wildfire in the State of California* (State of California, 2009).
59. Reineke, L. H. Perfecting a stand-density index for even-aged forests. *J. Agric. Res.* **46,** 627–638 (1933).
60. Smith, E. M., Larson, B. C., Kelty, M. J. & Ashton, P. M. S. *The Practice of Silviculture: Applied Forest Ecology* 9th edn (Wiley, 1996).
61. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **16,** 716–723 (1974).
62. Fox, J. & Weisberg, S. *An (R) Companion to Applied Regression* (Sage, 2011).
63. Lorenz, K. & Lal, R. *Carbon Sequestration in Forest Ecosystems* (Springer, 2010).
64. de Vries, W. *et al.* Ecologically implausible carbon response? *Nature* **451,** E1–E3 (2008).
65. Townsend, A. R., Braswell, B. H., Holland, E. A. & Penner, J. E. Spatial and temporal patterns in terrestrial carbon storage due to deposition of fossil fuel nitrogen. *Ecol. Appl.* **6,** 806–814 (1996).
66. Riebe, C. S., Kirchner, J. W. & Finkel, R. C. Sharp decrease in long-term chemical weathering rates along an altitudinal transect. *Earth Planet. Sci. Lett.* **218,** 421–434 (2004).
67. Lock, J., Kelsey, H., Furlong, K. & Woolace, A. Late Neogene and Quaternary landscape evolution of the northern California Coast Ranges: evidence for Mendocino Triple Junction tectonics. *Geol. Soc. Am. Bull.* 118, 1232–1246 10.1130/b25885.1 (2006).
68. Mitchell, C. E., Vincent, P., Weldon, R. J. & Richards, M. A. Present-day vertical deformation of the Cascadia Margin, Pacific Northwest, United States. *J. Geophys. Res. Solid Earth* **99,** 12257–12277 (1994).
69. Kelsey, H. M., Engebretson, D. C., Mitchell, C. E. & Ticknor, R. L. Topographic form of the Coast Ranges of the Cascadia Margin in relation to coastal uplift rates and plate subduction. *J. Geophys. Res. Solid Earth* **99,** 12245–12255 (1994).
70. Merritts, D. & Bull, W. B. Interpreting quaternary uplift rates at the Mendocino Triple Junction, northern California, from uplifted marine terraces. *Geology* **17,** 1020–1024 (1989).
71. Dahlgren, R. A. Geologic nitrogen as a source of soil acidity. *Soil Sci. Plant Nutr.* **51,** 719–723 (2005).
72. Casey, W. H. & Sposito, G. On the temperature-dependence of mineral dissolution rates. *Geochim. Cosmochim. Acta* **56,** 3825–3830 (1992).
73. Brady, P. V. & Carroll, S. A. Direct effects of CO_2 and temperature on silicate weathering—possible implications for climate control. *Geochim. Cosmochim. Acta* **58,** 1853–1856 (1994).
74. Verrall, R. E. Determination of activity coefficients for ammonium chloride at 25 °C. *J. Solution Chem.* **4,** 319–329 (1974).
75. Soil Survey Staff. *Soil Survey Geographic (SSURGO) Database for California* (United States Department of Agriculture, Natural Resources Conservation Service, 2010).
76. Beaudette, D. E. & O'Geen, A. T. *Algorithms for Quantitative Pedology: A Toolkit for Soil Scientists* (California Soil Resource Lab, Univ. of California, Davis, 2010).

LETTER

An earlier origin for the Acheulian

Christopher J. Lepre[1,2], Hélène Roche[3], Dennis V. Kent[1,2], Sonia Harmand[3], Rhonda L. Quinn[2,4], Jean-Philippe Brugal[5], Pierre-Jean Texier[6], Arnaud Lenoble[6] & Craig S. Feibel[2]

The Acheulian is one of the first defined prehistoric techno-complexes and is characterized by shaped bifacial stone tools[1–3]. It probably originated in Africa, spreading to Europe and Asia perhaps as early as ~1 million years (Myr) ago[4–6]. The origin of the Acheulian is thought to have closely coincided with major changes in human brain evolution, allowing for further technological developments[7,8]. Nonetheless, the emergence of the Acheulian remains unclear because well-dated sites older than 1.4 Myr ago are scarce. Here we report on the lithic assemblage and geological context for the Kokiselei 4 archaeological site from the Nachukui formation (West Turkana, Kenya) that bears characteristic early Acheulian tools and pushes the first appearance datum for this stone-age technology back to 1.76 Myr ago. Moreover, co-occurrence of Oldowan and Acheulian artefacts at the Kokiselei site complex indicates that the two technologies are not mutually exclusive time-successive components of an evolving cultural lineage, and suggests that the Acheulian was either imported from another location yet to be identified or originated from Oldowan hominins at this vicinity. In either case, the Acheulian did not accompany the first human dispersal from Africa[9,10] despite being available at the time. This may indicate that multiple groups of hominins distinguished by separate stone-tool-making behaviours and dispersal strategies coexisted in Africa at 1.76 Myr ago.

Sediments of the Nachukui formation exposed along the northwest shoreline of modern Lake Turkana in Kenya[11] (Fig. 1 and Supplementary Fig. 1) preserve several rich archaeological site complexes, among which is Kokiselei. Thus far, this complex is defined by ten sites, eight of which are found within a discrete 5-m interval of the middle part of the nearly 170-m-thick Kaitio member. Six of those sites have been tested or largely excavated but many other potential sites have been pinpointed. Most of these Kokiselei sites contain typical core/flake Oldowan assemblages[12]. Kokiselei 4 (KS4), however, holds an exceptional lithic assemblage that confirms the co-occurrence of the Oldowan and Acheulian at this site complex, indicating that the behavioural repertoire of early hominins in the area incorporated both technologies.

The KS4 assemblage (Supplementary Fig. 2) is characterized by the presence of pick-like tools with a trihedral or quadrangular section, unifacially or bifacially shaped crude hand-axes, and a few cores and flakes, all derived from the same mudstone bed. A single subsurface, *in situ* origin for KS4 is ensured by excavations at the main test trench that recovered several spectacular sets of refitted lithic artefacts (Supplementary Fig. 3). To the exception of a few cores made on basalt, the rest of the assemblage has been knapped from large cobbles or tabular clasts of locally available aphiric phonolite[13]. No vertebrate remains have been found within the test trenches, but the mudstone bed has yielded numerous non-human vertebrate fossils. The most frequent taxon is the large-sized hippopotamus *Hippopotamus gorgops* (adult and juvenile), but suids (*Kolpochoerus*, *Metridiochoerus*, *Notochoerus*?), rhinoceros (*Ceratotherium* sp.), equids (*Equus* and *Hipparion*) and a few carnivores (*Panthera*, Hyaenidae aff. *Crocuta*) are also present. Some isolated teeth and post-cranial elements represent bovids, and it is possible to identify Bovini (aff. *Syncerus*), Reduncini (*Kobus* sp.) and medium-sized Tragelaphini and Alcelaphini.

Broadly speaking, the Acheulian tools of KS4 come from a 15–20-m-thick interbedded series of gravels, sands and mudstones colloquially
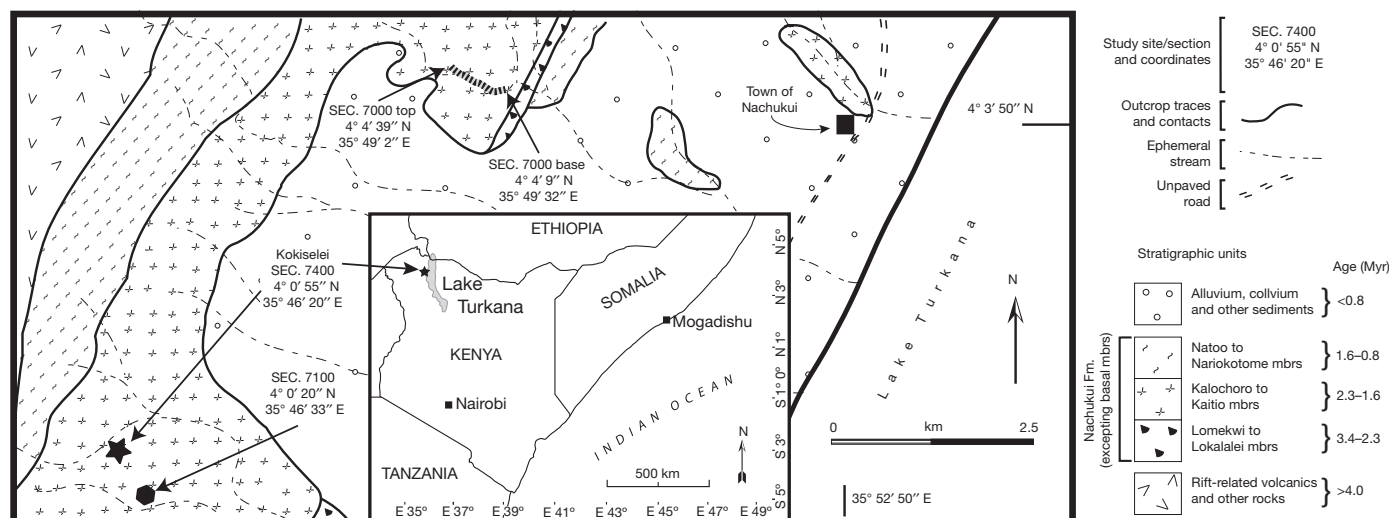
Figure 1 | Geological[11] and location map. Note the latitude and longitude coordinates provided for each place of investigation. Section 7000 (SEC. 7000) was surveyed ~200 m along the ephemeral stream; hence the indications of the base and top are given. The other two—SEC. 7100 and SEC. 7400 (the latter of which is where the Acheulian site KS4 was excavated)—derive from sheer outcrops, and the coordinates for only the top of the sections are given.

[1]Lamont-Doherty Earth Observatory, Palisades, New York 10964, USA. [2]Department of Earth and Planetary Sciences, Rutgers University, Piscataway, New Jersey 08854, USA. [3]UMR CNRS 7055, Université Paris Ouest Nanterre, 92023 Nanterre Cedex, France. [4]Department of Sociology, Anthropology, and Social Work, Seton Hall University, South Orange, New Jersey 07079, USA. [5]UMR CNRS 6636, Maison Méditerranéenne des Sciences de l'Homme, BP 647-F-13094, Aix-en-Provence Cedex 2, France. [6]UMR CNRS 5199-PACEA, Université de Bordeaux1, 33405 Talence, France.

referred to as the bird cliff beach complex (BCBC). Outcrops of the BCBC are part of a nearly continuous band of sediments belonging to the Kalochoro and Kaitio members that extends for over 7 km from northeast to southwest along the modern northwest shoreline of Lake Turkana, Kenya (Fig. 1). These members of the Nachukui formation record a transition from predominantly fluvial to lacustrine sedimentation in the Pleistocene Turkana basin during which the rest of East Africa is thought to have undergone ecosystem turnover in response to global climate forcing[14]. A first indication of lacustrine deposits occurs ~35 m above the base of the Kalochoro member, marked by the appearance of silty/clayey, often thinly bedded and finely laminated, ostracod-rich lacustrine mudstones, which coarsen upwardly into poorly sorted massive mudstones (Supplementary Figs 8–11). Thin lenticular units of gravel and sand become apparent near the top of the Kalochoro member, which heralds the appearance of the BCBC in the Kaitio member. Lithological units of the BCBC form metre-thick, coarsening-upward cycles of claystone, siltstone and sand and/or gravel. At KS4, the coarsest layers of the BCBC consist of gravelly sands that preserve abundant rhizoconcretions at their tops (Supplementary Figs 10 and 11). Claystones of the cycles contain thin lenses of mollusc shells and are dissected by slickensided fractures that define wedge-shaped aggregates of palaeosols. These claystones, as well

as the siltstone, can be very poorly sorted in places and include volcanic granules/pebbles. Such lithostratigraphic and sedimentary patterns suggest a dynamic environment along a palaeo-lakeshore.

In the Kokiselei region the erosive base of the BCBC occurs varyingly at 1–15 m above the Kaitio member's lowermost stratigraphic level, which is the KBS tuff[11] dated by $^{40}Ar/^{39}Ar$ to 1.869 ± 0.021 Myr ago[15]. The KBS tuff lies 78 m above the Kalochoro tuff, dated by $^{40}Ar/^{39}Ar$ to 2.331 ± 0.015 Myr ago[16], which defines the base of the Kalochoro member[11]. No direct numerical age constraints have been determined for the BCBC; however, regional geological mapping and tephrostratigraphic studies indicate that it is ~150 m below the base of the Lower Koobi Fora tuff[11] dated by $^{40}Ar/^{39}Ar$ to 1.476 ± 0.013 Myr ago[15]. Linear extrapolation from the Kalochoro and KBS tuffs and linear interpolation from the KBS and Lower Koobi Fora tuffs broadly constrain the KS4 Acheulian assemblage to between 1.72 and 1.81 Myr ago. Other dated tuffaceous beds between the KBS and Lower Koobi Fora tuffs in the Turkana basin that would further refine stratigraphic position have not been found in the Kokiselei region. To place the KS4 artefacts within a more constrained age context, we collected 148 orientated samples for palaeomagnetic analysis (see Supplementary Methods and Supplementary Figs 4–7) from sections of outcrop exposed at this archaeological site and adjacent locations (Fig. 1).
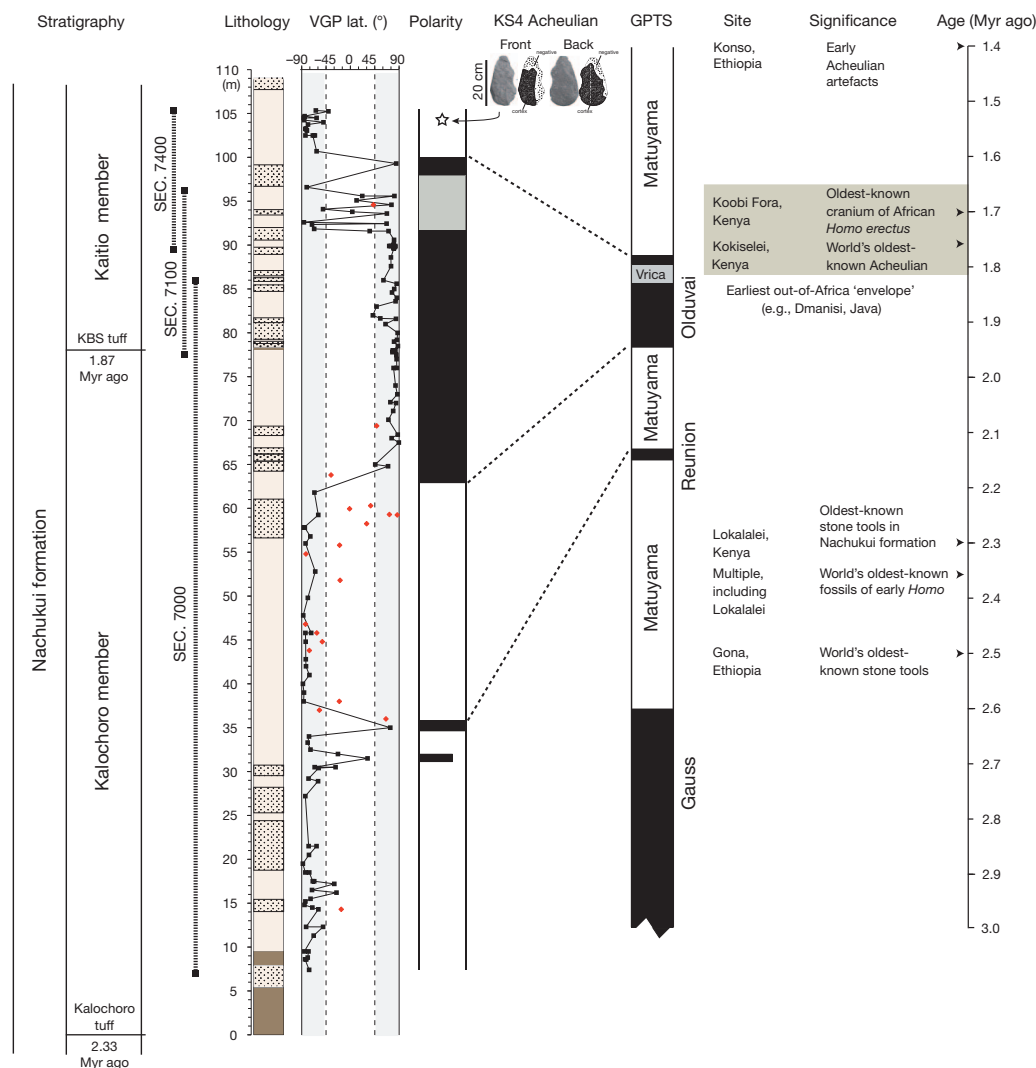
Figure 2 | Summary diagram. Left: members and tephrochronology[11,15,16], lithostratigraphy (brown, tuff/bentonite; buff, mudstone; small circles, sand; large circles, gravel) and virtual geomagnetic pole (VGP) latitudes (primary group (black symbols, $n = 129$) consists of reliable characteristic remanent magnetization directions; secondary group (red symbols, $n = 19$), unreliable) of the examined interval of the Nachukui formation. Right: the Early Stone Age in eastern Africa (refs 2, 3, 8, 12, 19, 25, 26 and this study) referenced to the geological ages of the oldest known out-of-Africa sites with hominin fossils[9,10] and the reversal chronology and age scale of the geomagnetic polarity timescale (GPTS)[17,23].

Our palaeomagnetic results allow us to recognize three main polarity intervals for the overall composite magnetostratigraphy from the examined interval for the Nachukui formation (Fig. 2). A lowest interval of almost entirely reverse polarity extends from 7 m to 63.5 m above the base of the Kalochoro member/tuff. The following interval of mostly normal polarity begins at 63.5 m above the base of the Kalochoro member/tuff and extends to approximately 100 m above the base. Overlying this long mostly normal magnetozone is a 5-m-thick interval of exclusively reverse polarity. $^{40}Ar/^{39}Ar$ dating of the Kalochoro tuff (~2.33 Myr ago) and the KBS tuff (~1.87 Myr ago) permits us to correlate accurately our magnetostratigraphy to the geomagnetic polarity timescale (GPTS) (Fig. 2). The polarity reversal at 63.5 m that is bracketed by these tuffs most probably represents the transition from the reverse Matuyama chron to the ensuing normal Olduvai subchron. Accordingly, the polarity reversal at ~100 m correlates with the subsequent transition from the normal Olduvai subchron to the ensuing part of the reverse Matuyama chron. This implies that the long, mostly normal magnetozone from 63.5 m to 100 m is the entire Olduvai subchron. Thus, the Olduvai in the Nachukui formation is evidently 36.5 m thick. Additional indications that we have located the extent of the Olduvai subchron in the Nachukui formation comes from the excellent agreement between the sedimentation rates for the overlapping intervals of the Kalochoro to KBS tuffs (~17 cm kyr^{-1}) and the identified base to top of the Olduvai subchron (~22 cm kyr^{-1}). If our outcrop sampling strategy was compromised by poor correlations, not enough vertical stratigraphic coverage, or unconformities, for example, then these two sets of independently derived sedimentation rates would be much more divergent, which is not the case—in fact, the experimental error associated with the radio-isotopic dates of the tuffs makes the two sets of sedimentation rates empirically indistinguishable.

At 28.5 and 32 m beneath the base of the Olduvai, our findings indicate the presence of two stratigraphic levels—each represented by one specimen—with positive inclinations and northerly virtual geomagnetic pole (VGP) latitudes that may correlate with the Reunion subchron. The positive inclination associated with the lower of the two levels might be the result of incomplete removal of a high-temperature magnetic component of specimen P022a carrier by hematite; however, the normal polarity associated with the upper level derives from well-resolved data of specimen 715-3a, which is more likely to represent the Reunion subchron (2.128–2.148 Myr ago[17]). Other work on the Turkana basin sequences and correlative deposits in southwest Ethiopia has documented the Reunion subchron occurring as two normal polarity intervals[18].

Our results reveal a complex, fine-scale pattern characterized by several short polarity excursions near the Olduvai to Matuyama boundary that is not unique to this particularly locality. It has been reported in this interval at other Turkana basin sequences[19], in oceanic cores[20,21] and at the former Plio-Pleistocene boundary and point stratotype section at Vrica from the eastern Mediterranean sapropel sequences[17,22,23]. The origins of the fine-scale structure characteristic of the top of the Olduvai subchron found at Turkana (ref. 19 and this study) and elsewhere[17,20–23] remain unclear, but might relate to the drop in the intensity of the geomagnetic field associated with the polarity reversal. The weak magnetic remanence of the sediment deposited in this low-intensity field may be more prone to resetting or overprinting in younger and comparatively stronger fields. Recent work suggests that such polarity excursions observed at Vrica could also reflect oxidation effects on the iron-bearing grains in the sapropel sequences[24]. Considering that post-Nachukui formation alluvial sediments mantle places of the landscape near KS4, some younger depositional or perhaps soil processes may have caused magneto-chemical alterations that might contribute to some of the complexity. Nevertheless, the similar palaeomagnetic reversal character for the end Olduvai at sites in different global geographic settings offers an exceptional correlation tool and provides additional support for our magnetostratigraphic interpretations. The stratigraphic position of

KS4 is 4.5 m above the Olduvai to Matuyama boundary. An age model based on a cubic spline fit to six chronostratigraphic tie-points (Kalochoro tuff, Reunion subchron, base of the Olduvai subchron, KBS tuff, top of the Olduvai subchron, and Lower Koobi Fora tuff) provides an estimate of 1.76 Myr ago for the KS4 Acheulian assemblage (Fig. 3, see also Methods). This is more than 350 kyr older than the early Acheulian artefacts from Konso, Ethiopia[3].

An origin for the Acheulian back to 1.76 Myr ago is close in age to partial cranium KNM-ER 3733[19], which is ostensibly the most definitive evidence for the antiquity of African *Homo erectus sensu lato*, considering it is arguably a more anatomically diagnostic specimen, and thus better understood taxonomically as compared to possible conspecific fossils older than 1.7 Myr[25]. Several hypotheses link the development of the Acheulian with the initial evolution of *H. erectus*. This is mainly because of a similar geographic origin for the two[8], and the large-brained species persisted when many, if not all other, *Homo* taxa went extinct[26] during the evolution of the characteristically made Acheulian post-1.5-Myr ago. An earlier Acheulian origin, coeval with sympatric *Homo* species[26], strengthens the possibility that more than one tool-making hominin existed at 2.0–1.5 Myr ago.

Homo erectus is traditionally thought to be the first hominin to disperse from Africa, yet the oldest known out-of-Africa fossil hominin sites lack stone tools or preserve only Oldowan-style artefacts[9,10]. If indeed the first out-of-Africa hominin possessed Acheulian technology, then it is expected that evidence of this techno-culture should also be found dispersed throughout the Old World. However, archaeological sites older than ~1 Myr preserving the Acheulian are not abundantly documented from the Middle East, Europe or Asia, and are younger than the oldest known out-of-Africa hominin fossil localities dated at 1.7 Myr ago[4–6,9,10]. Our data indicate that the earliest development of the Acheulian occurred in Africa at 1.76 Myr ago and was contemporaneous with or perhaps pre-dated the earliest hominin dispersals into Eurasia. Yet, the difference between the ages for the oldest known Acheulian artefacts in the world from Africa and the oldest known Acheulian artefacts from Eurasia raises the likelihood that the first Eurasian hominins derived from an African population lacking Acheulian culture. Potentially, two hominin groups coexisted in Africa at 1.76 Myr ago. One of these groups could have developed the
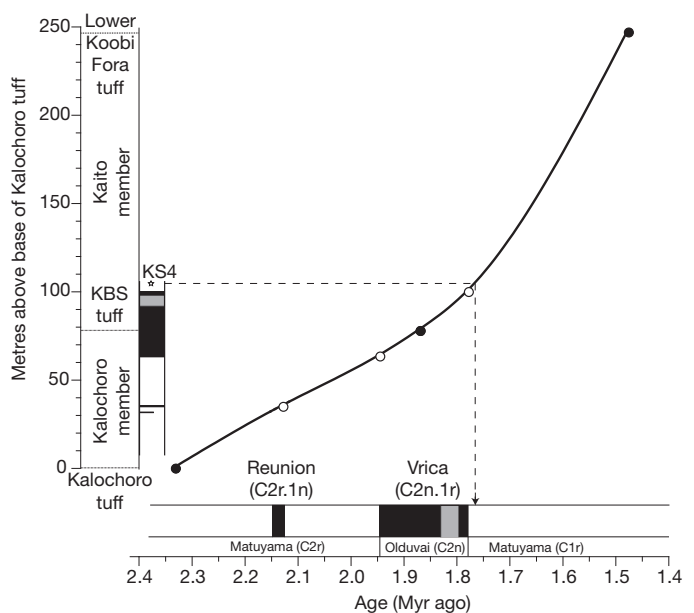
Figure 3 | Age model. Cubic spline curve fitted to stratigraphic levels above the Kalochoro tuff versus age of magnetostratigraphic subchrons[17,23] (open circles) and dated tuffs[15,16] (filled circles); dashed line shows how the stratigraphic level of KS4 was used to derive its age (1.76 Myr ago) from the spline curve.

Acheulian technology but remained in Africa. The other could have lacked the cognitive ability and/or technological knowledge to manufacture the Acheulian technology and did not carry it into Eurasia. This division may indicate different behavioural aptitudes for separate African species (for example, *H. erectus sensu lato* versus *Homo habilis sensu lato*) or a within-species cultural disparity. In any event, it seems that a second hominin dispersal with Acheulian technology or a diffusion of this technology took place later, leading to the widespread occurrence of this Early Stone Age tradition in the circum-Mediterranean area and elsewhere after ~1 Myr ago[4,5].

METHODS SUMMARY

Orientations by magnetic compass and clinometer of planar faces were marked before removing hand-cut blocks from outcrops. Samples were taken at one-metre intervals or as the occurrence of fine-grained strata permitted. At least one independent block sample was taken from each interval resulting in 148 independent samples, from which one or more specimens were cut for processing. Magnetic remanence measurements were made with a 2G Model 760 DC-SQUID rock magnetometer in the shielded room of the Paleomagnetics Laboratory at Lamont-Doherty Earth Observatory. Natural remanent magnetizations of all sample specimens were subjected to progressive thermal demagnetization using an initial step of 100 °C, seven steps at 50 °C increments to 450 °C, and five steps at 25 °C increments to 575 °C. Magnetic susceptibility values were determined with a Bartington MS2B instrument initially and after each heating step to monitor for magneto-chemical alteration. Virtual geomagnetic pole (VGP) latitudes were calculated from the characteristic remanent magnetization (ChRM) directions determined from principal component analysis[27] and Zijderveld demagnetization diagrams[28] (Supplementary Table 1). Reliable ChRM directions are characterized by maximum angular deviation (MAD) values of less than 15° of a component that linearly converges towards the origin over five high-temperature steps. VGP latitude for each specimen was plotted in stratigraphic position to determine magnetostratigraphy.

The numerical age of KS4 was estimated through a model of the rate of sediment accumulation for the deposits encasing the site using a cubic spline function through data for the Kalochoro tuff (0 m, 2.331 Myr ago[16]), Reunion subchron (35 m, midpoint 2.138 Myr ago[17,23]), base of Olduvai subchron (63.5 m, 1.945 Myr ago[17,23]), KBS tuff (78 m, 1.869 Myr ago[15]), top of Olduvai subchron (100 m, 1.778 Myr ago[17,23]), and Lower Koobi Fora tuff (247 m, 1.476 Myr ago[15]). This method places the KS4 Acheulian assemblage (104.5 m) at 1.76 Myr ago.

1. Leakey, M. *Olduvai Gorge: Excavations in Beds I & II* 1960–1963 (Cambridge Univ. Press, 1971).
2. Isaac, G., Ll. & Curtis, G. H. Age of early Acheulean industries from the Peninj Group, Tanzania. *Nature* **249,** 624–627 (1974).
3. Asfaw, B. et al. The earliest Acheulean from Konso-Gardula. *Nature* **360,** 732–735 (1992).
4. Goren-Inbar, N. et al. Pleistocene milestones on the out-of-Africa corridor at Gesher Benot Ya'aqov, Israel. *Science* **289,** 944–947 (2000).
5. Scott, G. R. & Gibert, L. The oldest handaxes in Europe. *Nature* **461,** 82–85 (2009).
6. Pappu, S. et al. Early Pleistocene presence of Acheulian hominins in south India. *Science* **331,** 1596–1599 (2011).
7. Gowlett, J. A. J. in *Stone Age Prehistory* (eds Bailey, G. N. and Callow, O.) 243–260 (Cambridge Univ. Press, 1986).
8. Klein, R. *The Human Career: Human Biological and Cultural Origins* 3rd edn (University Chicago Press, 2009).
9. Swisher, C. C. III, Curtis, G. H., Jacob, T., Getty, A. G. & Suprijo, A. Age of the earliest known hominids in Java, Indonesia. *Science* **263,** 1118–1121 (1994).
10. Gabunia, L. et al. Earliest Pleistocene hominid cranial remains from Dmanisi, Republic of Georgia: taxonomy, geological setting, and age. *Science* **288,** 1019–1025 (2000).
11. Harris, J. M., Brown, F. H. & Leakey, M. G. Geology and paleontology of Plio-Pleistocene localities west of Lake Turkana, Kenya. *Contrib. Sci.* **399,** 1–128 (1988).
12. Roche, H. et al. Les sites archéologiques plio-pléistocènes de la formation de Nachukui, Ouest-Turkana, Kenya: bilan synthétique 1997–2001. *C. R. Palevol* **2,** 663–673 (2003).
13. Harmand, S. Raw material and economic behaviours at Oldowan and Acheulean sites in the West Turkana region, Kenya. In *Lithic Materials and Paleolithic Societies* (eds Adams, B. & Blades, B.). 3–14 (Wiley-Blackwell, 2009).
14. Feibel, C. S., Harris, J. M. & Brown, F. H. in *Koobi Fora Research Project* Vol. 3 *The Fossil Ungulates: Geology, Fossil Artiodactyls, and Palaeoenvironments* (ed. Harris, J. M.) 321–370 (Clarendon, 1991).
15. McDougall, I. & Brown, F. H. Precise 40Ar/39Ar geochronology for the upper Koobi Fora Formation, Turkana Basin, northern Kenya. *J. Geol. Soc. Lond.* **163,** 205–220 (2006).
16. McDougall, I. & Brown, F. H. Geochronology of the pre-KBS Tuff sequence, Omo Group, Turkana Basin. *J. Geol. Soc. Lond.* **165,** 549–562 (2008).
17. Lourens, L., Hilgen, F., Shackleton, N. J., Laskar, J. & Wilson, D. in *A Geologic Time Scale* (eds Gradstein, F., Ogg, J. & Smith, A.) 409–440 (Cambridge Univ. Press, 2004).
18. Kidane, T., Otofuji, Y.-I., Brown, F. H., Takemoto, K. & Eshete, G. Two normal paleomagnetic polarity intervals in the lower Matuyama Chron recorded in the Shungura Formation (Omo Valley, Southwest Ethiopia). *Earth Planet. Sci. Lett.* **262,** 240–256 (2007).
19. Lepre, C. J. & Kent, D. V. New magnetostratigraphy for the Olduvai Subchron in the Koobi Fora Formation, northwest Kenya, with implications for early Homo. *Earth Planet. Sci. Lett.* **290,** 362–374 (2010).
20. Ninkovich, D., Opdyke, N. D., Heezen, B. C. & Foster, J. H. Paleomagnetic stratigraphy, rates of deposition and tephrachronology in North Pacific deep-sea sediments. *Earth Planet. Sci. Lett.* **1,** 476–492 (1966).
21. Clement, B. M. & Kent, D. V. Geomagnetic polarity transition records from five hydraulic piston core sites in the North Atlantic. *Initial Rep. Deep Sea Drill. Proj.* **94,** 831–852 (1987).
22. Tauxe, L., Opdyke, N. D., Pasini, G. & Elmi, C. Age of the Plio-Pleistocene boundary in the Vrica section, southern Italy. *Nature* **304,** 125–129 (1983).
23. Zijderveld, J. D. A., Hilgen, F. J., Langereis, C. G., Verhallen, P. J. J. M. & Zachariasse, W. J. Integrated magnetostratigraphy and biostratigraphy of the upper Pliocene–lower Pleistocene from the Monte Singa and Crotone areas in Calabria, Italy. *Earth Planet. Sci. Lett.* **107,** 697–714 (1991).
24. Roberts, A. P., Florindo, F., Larrasoaña, J. C., O'Regan, M. A. & Zhao, X. Complex polarity pattern at the former Plio–Pleistocene global stratotype section at Vrica (Italy): Remagnetization by magnetic iron sulphides. *Earth Planet. Sci. Lett.* **292,** 98–111 (2010).
25. Wood, B. A. *Koobi Fora Research Project* Vol. 4 *Hominid Cranial Remains from Koobi Fora* (Clarendon, Oxford, 1991).
26. Spoor, F. et al. Implications of new early *Homo* fossils from Ileret, east of Lake Turkana, Kenya. *Nature* **448,** 688–691 (2007).
27. Kirschvink, J. L. The least-squares line and plane and the analysis of palaeomagnetic data. *Geophys. J. R. Astron. Soc.* **62,** 699–718 (1980).
28. Zijderveld, J. D. A. A. C. in *Methods in Paleomagnetism* (eds Collinson, D. W., Creer, K. M. & Runcorn, S. K.) 254–286 (Elsevier, 1967).

LETTER

Transient dynamics of an altered large marine ecosystem

Kenneth T. Frank[1], Brian Petrie[1], Jonathan A. D. Fisher[2] & William C. Leggett[2]

Overfishing of large-bodied benthic fishes and their subsequent population collapses on the Scotian Shelf of Canada's east coast[1,2] and elsewhere[3,4] resulted in restructuring of entire food webs now dominated by planktivorous, forage fish species and macroinvertebrates. Despite the imposition of strict management measures in force since the early 1990s, the Scotian Shelf ecosystem has not reverted back to its former structure. Here we provide evidence of the transient nature of this ecosystem and its current return path towards benthic fish species domination. The prolonged duration of the altered food web, and its current recovery, was and is being governed by the oscillatory, runaway consumption dynamics of the forage fish complex. These erupting forage species, which reached biomass levels 900% greater than those prevalent during the pre-collapse years of large benthic predators, are now in decline, having outstripped their zooplankton food supply. This dampening, and the associated reduction in the intensity of predation, was accompanied by lagged increases in species abundances at both lower and higher trophic levels, first witnessed in zooplankton and then in large-bodied predators, all consistent with a return towards the earlier ecosystem structure. We conclude that the reversibility of perturbed ecosystems can occur and that this bodes well for other collapsed fisheries.

The recent demonstration that overfishing of large-bodied predators in the northwest Atlantic initiated a trophic cascade, typified by reciprocal changes in biomass between adjacent trophic levels extending to the base of the food web[1,2], overturned the long-held view that large marine ecosystems are resistant to restructuring[5]. It has been proposed[6,7] that such trophic cascades are characteristic of ecosystems that have been transformed into undesirable states involving large changes in ecological functions and/or economic resources[8,9]. Although the excessive consumption characteristic of trophic cascades may be unstable[6], whether, how, and on what time scales such altered, diverse food webs and their key species and functional groups will recover remains unknown[1,3,4]. This has led to controversy regarding the efficacy of and experimentation with strategies based on conventional management approaches such as moratoria on exploitation, culling and re-stocking intended to return ecosystems to their former structure[10–13]. Using four decades of high quality, annual, fishery-independent data (see Methods) representative of multiple trophic levels on the eastern Scotian Shelf (Supplementary Fig. 1), we document the transient nature of its altered ecosystem and its return towards dominance by large-bodied predators.

The collapse of the northwest Atlantic cod (*Gadus morhua*) and several other large predatory fishes in the early 1990s (Fig. 1a), caused principally by over-fishing[14,15], precipitated the first documented open ocean trophic cascade in a large marine ecosystem[1]. The total biomass of cod, one of the ecosystem's dominant species, has hovered at less than 5% of pre-collapse levels for almost two decades despite the implementation of strict regulations forbidding their capture[15].

Recent investigations[13,16] have provided strong evidence that, following these collapses, the eastern Scotian Shelf, and other northwest Atlantic ecosystems in which similar collapses occurred, moved to

apparent alternate states in which planktivorous forage fishes and macroinvertebrates became the dominant predators[3]. Released from predation on the eastern Scotian Shelf, the biomass of forage fishes increased by 900% (Fig. 1b) and macroinvertebrates by 200% compared to the pre-collapse years[13]. They then competed directly with and/or preyed upon the early life stages of their once benthic predators, a phenomenon termed predator–prey reversal[17] which seems to be one of the leading causes of the delayed recovery of the benthic fish complex in this and other large marine ecosystems[4,17,18]. Although forage fish constitute approximately half the diet of an expanding, resident grey seal (*Halichoerus grypus*) population, estimates of their consumption of pelagic fish species (1995–2000) were only 35% of the benthic
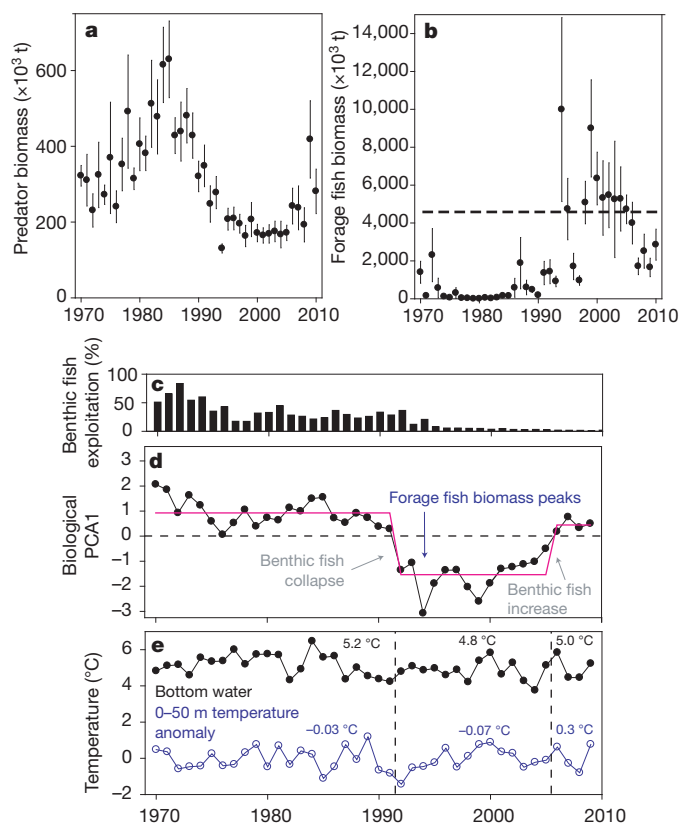
Figure 1 | Variability of the eastern Scotian Shelf ecosystem. a–e, Data (Supplementary Fig. 1) based on large-bodied benthic fish (**a**), their prey (forage fishes, **b**) with estimated carrying capacity (dashed line), benthic fish exploitation history expressed as annual per cent biomass removal (**c**), changing ecosystem structure based on the leading mode (PCA1) of biotic data spanning four trophic levels and the demarcation of regimes[29] of 22, 14 and 4 years duration (pink solid line) (**d**), and temperatures with averages shown for the three regimes (dashed vertical lines) (**e**). Vertical bars in **a** and **b** show ±s.e.m. (*n* = 27).

[1]Ocean Sciences Division, Bedford Institute of Oceanography, P.O. Box 1006, Dartmouth, Nova Scotia B2Y 4A2, Canada. [2]Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada.

fish complex[13] and insufficient to suppress the outbreaks and biomass variability of forage fishes (Fig. 1b).

The changing status of the eastern Scotian Shelf ecosystem exhibited two transitions. A period of intensive fishing, when aggregate landings of benthic fishes averaged close to 105,000 tonnes (t) representing annual biomass removals of >50% (Fig. 1c), resulted in the first transition centred in 1991–1992 from a 'pre-collapse' state dominated by benthic fish species to a 'collapsed' state dominated by forage fish species[1]. A cod and haddock fishing moratorium implemented in 1993 had the desired effect of reducing aggregate exploitation (<5% since 2000; Fig. 1c) but did not produce the anticipated recovery (Supplementary Figs 2 and 3). The second transition, centred in 2005–2006, represents a return towards a 'recovering' state of benthic fish domination described here (Fig. 1d and Methods). An additional bottom trawl survey, beginning in 1986, revealed a similar pattern of collapse and recent increase in benthic fish biomass (Supplementary Fig. 4).

The physical environment during the three states, assessed using annual bottom temperatures, 0–50 m water temperatures (Fig. 1e) and water column stratification (Supplementary Information), showed only minor changes. The bottom temperatures during the collapsed state were only 0.33 °C and 0.24 °C lower than during the pre-collapse and recovering periods, respectively. The magnitude of this temperature change would have only minimal or no effect on individual and population growth rates as well as other life history traits (Methods). Further, the dominant large-scale atmospheric forcing mechanism in the western North Atlantic (that is, the North Atlantic Oscillation) induces a bimodal response of ocean temperatures with a nodal point in bottom temperature occurring in the middle of the eastern Scotian Shelf[19]. Consequently, the temperature response to such forcing in this region is minimal and is reflected in the dampening of regional variability in other biological properties such as species richness[20]. Differences in water column temperature anomalies were also slight for the three periods: on average, temperatures during the pre-collapse and collapsed periods were within 0.1 °C, during the forage fish outbreak temperatures were elevated by about 0.4 °C; overall, temperatures varied over a range of about 2 °C. There was no relationship between water column temperatures and forage fish biomass at zero lag (Supplementary Fig. 5; correlation coefficient, $r = 0.02$) or for lags (forage fish biomass relative to temperature) up to four years (Methods). The minor increases in stratification occurred primarily during summer, outside of the peak period of phytoplankton production (Methods). The timing and magnitude of this ongoing recovery of the benthic fish complex was initiated and is being sustained by naturally induced changes in the dynamics of their former prey, and the resulting impact on the total ecosystem, more so than by external climatic influences (Fig. 1e).

The second and most recent ecosystem transition began with a marked decline in the biomass of the unfished, forage fish complex dominated by northern sand lance (*Ammodytes dubius*), capelin (*Mallotus villosus*), and Atlantic herring (*Clupea harengus*). The aggregate biomass of these species peaked in 1994 and 1999 at approximately 10 million t, which exceeded the estimated carrying capacity of 4.3 million t for the eastern Scotian Shelf ecosystem (Fig. 1b; Methods). Subsequently, their total biomass rapidly declined at an average rate of 0.5 million t per year to current levels near 3 million t. Such eruptions followed by crashes involving fast growing, highly opportunistic species are known to occur in other ecosystems freed from predatory control[21,22].

Physiological changes and cascading food web effects associated with the overshoot of the pelagic forage fish were evident. Relative weights, an index of physiological condition available since 1970, of the three dominant forage fish species showed coherent changes (31 out of 41 years with same sign) throughout the entire period with relative high condition indices from the 1970s to the early 1990s followed by sustained declines beginning about 1994 (Fig. 2a). This points to the pelagic species having inadequate food resources at
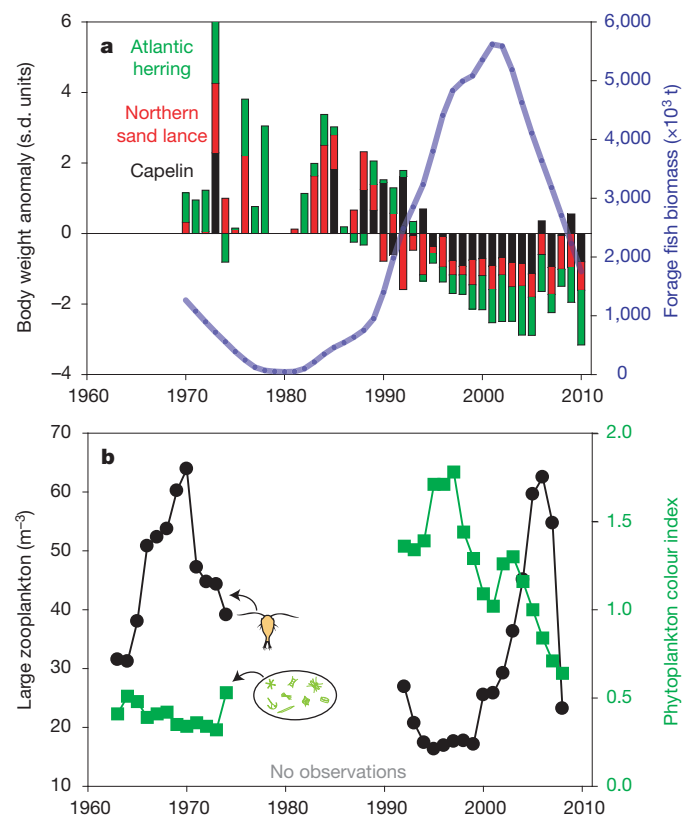
Figure 2 | Physiological changes in forage fish species and resultant food web effects. a, Species-specific body weight anomalies (stacked histograms) and smoothed forage fish biomass (solid line; 30% LOESS; Supplementary Fig. 8). **b**, Time series of large-bodied zooplankton abundance and phytoplankton colour index (a measure of abundance) from the Continuous Plankton Recorder survey (http://www.safhos.ac.uk). Recent (2000–2007) increases in large-bodied zooplankton, coincident with major reductions in forage fish biomass shown in **a**, indicate a weakening of the inhibitory effects of the predator–prey reversal mechanism to benthic fish species recovery.

increased abundances. The density of large-bodied zooplankton, which has varied inversely ($r = -0.32$, $n = 22$ years) with the forage fish biomass (their principal predators), reached a broad minimum of about 17 individuals per m³ at the approximate peak (1994) of the forage group biomass—a signature of excessive grazing. Large-bodied zooplankton species increased rapidly (1997–2007) by a factor of four in six to seven years to levels observed before the forage fish eruption, the final year in the series being an exception (Fig. 2b). Moreover, the standing stock of phytoplankton, which has varied inversely with the abundance of large-bodied zooplankton ($r = -0.71$), declined by 40% during the ongoing recovery of the benthic fish complex. These reciprocal relationships between adjacent trophic levels are consistent with the trophic cascade model[1].

Most revealing is the fact that the trajectory of forage fish biomass changes following the benthic fish collapse approximates a damped harmonic oscillator (Fig. 3), providing key biological insights from the derived parameter values including estimates of the period, 5 years (the approximate life cycle of this functional group), and dampening time, 7 years. The 7-year dampening represented a 78% decrease in forage fish biomass between 1994, the time of the peak amplitude, and 2005, the end of the regime dominated by this group (Fig. 1d). Studies of terrestrial herbivores indicate that the cycle period is strongly dependent on body size[22]; based on the range of body sizes of the forage fishes we studied (weight range: 0.04–0.3 kg), the cycle period is similar to those estimated from outbreaks of similar-sized terrestrial species (3.6 to 5.9 years)[22]. This dynamic, oscillatory behaviour indicates that the internal damping capacity of the ecosystem, and not solely management
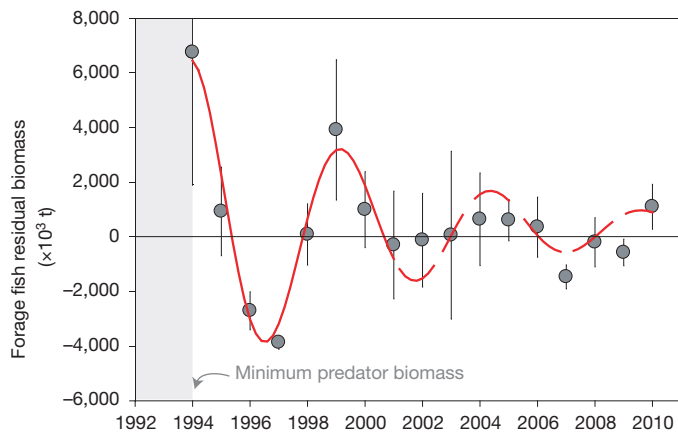
Figure 3 | Post-perturbation forage fish variability. The variability (solid points ± s.e.m., $n = 27$), described by a damped harmonic oscillator fitting the 1994–2001 observations (solid line), gives an amplitude of $6,100 \times 10^3$ t, dampening rate = 6.9 years, and oscillation period = 5.2 years (Methods). Shaded area indicates period of greatest population growth. Aggregate prey biomass is expressed as an anomaly from the 30% LOESS filter (Supplementary Fig. 8). This solution is extrapolated to the end of the record (broken line). In theory[30], a population or functional group that is characterized by damped oscillations will overshoot and then undershoot the carrying capacity.

strategies, probably initiated the return of the eastern Scotian Shelf ecosystem towards its former structure and the restoration of food web stability.

We propose that this crash in forage fish biomass led to a reduction in the intensity of predator–prey reversal which, combined with increased food availability to the benthic fish larval stages resulting from the corresponding and related growth in the abundance of large zooplankton, provided a window of opportunity for the recovery of the once dominant benthic predators. Since 2006 the benthic fish biomass has attained levels approaching those observed during the pre-collapse period (Fig. 1a). Atlantic cod and redfish (*Sebastes* spp.) have reached levels not seen since the early 1990s and haddock (*Melanogrammus aeglefinus*) to an unprecedented high (Supplementary Fig. 6). Enhanced recruitment (by a factor of 5.3, recovering/pre-collapse; Supplementary Fig. 7) and improvements in post-recruitment survivorship, which for cod and haddock has increased by factors of 12 and 70 times compared to the collapsed period, contributed to these changes (Supplementary Table 1). In addition, three of the four benthic fish species which are routinely aged (cod, pollock and silver hake) have shown 8–18% increases in average weight at age during 2006–2010 relative to the 1992–2005 post-collapse period (Supplementary Table 2).

The generally positive response of the large-bodied zooplankton to the declining forage fish biomass supports this hypothesis. The early life stages of most of the benthic fish species form an integral part of the large-bodied zooplankton complex and their survival and contribution to recruitment would have benefited from the same forces that have led to the increase in large-bodied zooplankton abundance noted above. Increased predation by the expanding benthic fish complex on the forage fish community should accelerate this trend.

This unfolding drama held many surprises, including the prolonged recovery of the benthic predator complex, despite the moratorium on directed fishing for cod and haddock, the establishment of a closed area on the western offshore banks that preceded the fishing moratorium[23], and the promotion of new and experimental fisheries[24] to divert fishing effort away from the traditional species. Although the current trajectory is positive, several factors could alter ongoing ecosystem recovery. The current high levels of recruitment and survivorship of the benthic fish complex, if sustained, could accelerate the recovery. The current dominance of haddock over cod also raises the question of whether the species makeup of the ecosystem will return to that which prevailed before the collapse. Furthermore, recovery in other over-exploited

ecosystems such as the Black Sea, Northern Benguela, the Sea of Japan, and elsewhere has been delayed by jellyfish blooms[25], the presence of invasive species and by eutrophication[3], all of which are possible in the system we describe. The widespread body size reductions of benthic fishes documented for other exploited northwest Atlantic systems[26,27], if not reversed, could also slow the recovery of the benthic fish complex and adversely affect food web structure[26]. The evolving global climate could alter the ecosystem positively or negatively.

These uncertainties notwithstanding, the answer to the critical question of whether or not such profound changes in the dynamics of large marine ecosystems are reversible seems to be 'yes'. This bodes well for other perturbed, formerly cod-dominated systems at latitudes to the north of the eastern Scotian Shelf that have yet to recover. Indeed, subtle signs of cod recovery have been appearing in other sub-arctic northwest Atlantic ecosystems during the past few years[28]. However, the time scales for their recovery are likely to be greater given the lower water temperatures (equates to slower turnover times) and their reduced species richness and, for some, because of the continued exploitation of cod and other large-bodied benthic fish species[13,15,18].

METHODS SUMMARY

Annual fishery-independent, randomly stratified bottom trawl surveys on the continental shelf off eastern Nova Scotia (1970–2010) during July–August provided biomass and variance estimates for functional groupings of fifteen, commercially exploited benthic and three forage fish species. A March survey (1986–2010) was used to assess the benthic fish dynamics further. Benthic fish exploitation levels were expressed as the ratio of landings to survey-estimated biomass. Ageing data, available for four benthic fish species, were used to calculate growth and mortality rates. Correction factors were applied to the typically under-sampled forage species. A 30% linear LOESS filter was applied to the forage fish biomass data to isolate high-frequency variability; the resultant anomalies were least squares fit to a damped harmonic oscillator equation (Methods). We quantified the time-averaged ecosystem carrying capacity of forage fishes using seasonally averaged, zooplankton data and production to biomass ratios obtained from the literature (Methods). Annual anomalies of the individual average body weights (total biomass/total abundance), an index of physiological condition, were derived for each forage species. Lower trophic level data (phytoplankton colour index and zooplankton abundance time series) were obtained from the Continuous Plankton Recorder survey (http://www.safhos.ac.uk). Zooplankton were grouped into large (≥ 2 mm carapace length) and small (<2 mm) species. The eastern Scotian Shelf ecosystem status was assessed by principal components analysis of the five biological time series. To determine whether and when ecosystem transitions occurred, the dominant principal component axis (PCA1) was subjected to a sequential t-test analysis of the regime shift method (STARS)[29], which identifies the magnitude and direction of significant shifts. Temperature, salinity and density observations were obtained from directed and opportunistic ship-based sampling.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Frank, K. T., Petrie, B., Choi, J. S. & Leggett, W. C. Trophic cascades in a formerly cod-dominated ecosystem. *Science* **308**, 1621–1623 (2005).
2. Frank, K. T., Petrie, B., Shackell, N. L. & Choi, J. S. Reconciling differences in trophic control in mid-latitude marine ecosystems. *Ecol. Lett.* **9**, 1096–1105 (2006).
3. Daskalov, G. M., Grishin, A. N., Rodionov, S. & Mihneva, V. Trophic cascades triggered by overfishing reveal possible mechanisms of ecosystem regime shifts. *Proc. Natl Acad. Sci. USA* **104**, 10518–10523 (2007).
4. Casini, M. *et al.* Multi-level trophic cascades in a heavily exploited open marine ecosystem. *Proc. R. Soc. Lond. B* **275**, 1793–1801 (2008).
5. Steele, J. H. & Collie, J. S. in *The Global Coastal Ocean: Multiscale Interdisciplinary Processes* (eds Robinson, A. R. & Brink, K.) vol. 13, chap. 21 (Harvard, 2004).
6. Strong, D. R. Are trophic cascades all wet? Differentiation and donor-control in speciose ecosystems. *Ecology* **73**, 747–754 (1992).
7. Pace, M. L., Cole, J. J., Carpenter, S. R. & Kitchell, J. F. Trophic cascades revealed in diverse ecosystems. *Trends Ecol. Evol.* **14**, 483–488 (1999).
8. Jackson, J. B. C. *et al.* Historical overfishing and the recent collapse of coastal ecosystems. *Science* **293**, 629–638 (2001).
9. Myers, R. A., Baum, J. K., Shepherd, T. D., Powers, S. P. & Peterson, C. H. Cascading effects of the loss of apex predatory sharks from a coastal ocean. *Science* **315**, 1846–1850 (2007).
10. Persson, L. *et al.* Culling prey promotes predator recovery—alternative states in a whole-lake experiment. *Science* **316**, 1743–1746 (2007).

11. Terborgh, J. & Estes, J. A. *Trophic Cascades: Predators, Prey, and the Changing Dynamics of Nature* (Island, 2010).
12. Yodzis, P. Must top predators be culled for the sake of fisheries? *Trends Ecol. Evol.* **16,** 78–84 (2001).
13. Bundy, A., Heymans, J. J., Morisette, L. & Savenkoff, C. Seals, cod and forage fish: a comparative exploration of variations in the theme of stock collapse and ecosystem change in four Northwest Atlantic ecosystems. *Prog. Oceanogr.* **81,** 188–206 (2009).
14. Myers, R. A. & Worm, B. Extinction, survival or recovery of large predatory fishes. *Phil. Trans. R. Soc. B* **360,** 13–20 (2005).
15. Shelton, P. A. *et al.* Fishing under low productivity conditions is further delaying recovery of Northwest Atlantic cod (*Gadus morhua*). *Can. J. Fish. Aquat. Sci.* **63,** 235–238 (2006).
16. Savenkoff, C. *et al.* Effects of fishing and predation in a heavily exploited ecosystem: comparing periods before and after the collapse of groundfish in the southern Gulf of St. Lawrence (Canada). *Ecol. Modell.* **204,** 115–128 (2007).
17. Fauchald, P. Predator–prey reversal: a possible mechanism for ecosystem hysteresis in the North Sea? *Ecology* **91,** 2191–2197 (2010).
18. Petrie, B., Frank, K. T., Shackell, N. L. & Leggett, W. C. Structure and stability in exploited marine fish communities: quantifying critical transitions. *Fish. Oceanogr.* **18,** 83–101 (2009).
19. Petrie, B. Does the North Atlantic Oscillation affect hydrographic properties on the Canadian Atlantic continental shelf? *Atm.-Ocean* **45,** 141–151 (2007).
20. Fisher, J. A. D., Frank, K. T., Petrie, B., Leggett, W. C. & Shackell, N. L. Temporal dynamics within a contemporary latitudinal diversity gradient. *Ecol. Lett.* **11,** 883–897 (2008).
21. Forsyth, D. M. & Caley, P. Testing the irruptive paradigm of large-herbivore dynamics. *Ecology* **87,** 297–303 (2006).
22. Peterson, R. O., Page, R. E. & Dodge, K. M. Wolves, moose, and the allometry of population cycles. *Science* **224,** 1350–1352 (1984).
23. Frank, K. T., Shackell, N. L. & Simon, J. E. An evaluation of the Emerald/Western Bank juvenile haddock closed area. *ICES J. Mar. Sci.* **57,** 1023–1034 (2000).
24. Anderson, S. C., Lotze, H. K. & Shackell, N. L. Evaluating the knowledge base for expanding low-trophic-level fisheries in Atlantic Canada. *Can. J. Fish. Aquat. Sci.* **65,** 2553–2571 (2008).
25. Richardson, A. J., Bakun, A., Hays, G. C. & Gibbons, M. J. The jellyfish joyride: causes, consequences and management responses to a more gelatinous future. *Trends Ecol. Evol.* **24,** 312–322 (2009).
26. Shackell, N. L., Frank, K. T., Fisher, J. A. D., Petrie, B. & Leggett, W. C. Decline in top predator body size and changing climate alter trophic structure in an oceanic ecosystem. *Proc. R. Soc. B* **277,** 1353–1360 (2010).
27. Fisher, J. A. D., Frank, K. T. & Leggett, W. C. Breaking Bergmann's rule: truncation of Northwest Atlantic marine fish body sizes. *Ecology* **91,** 2499–2505 (2010).
28. Murawski, S. A. Rebuilding depleted fish stocks: the good, the bad, and mostly, the ugly. *ICES J. Mar. Sci.* **67,** 1830–1840 (2010).
29. Rodionov, S. N. A sequential algorithm for testing climate regime shifts. *Geophys. Res. Lett.* **31,** L09204 (2004).
30. Rooney, N., McCann, K., Gellner, G. & Moore, J. C. Structural asymmetry and the stability of diverse food webs. *Nature* **442,** 265–269 (2006).

METHODS

Large-bodied benthic predators, aggregate forage fish biomass, zooplankton functional groups and chlorophyll. The primary data used in the analyses were from July–August surveys, initiated in 1970, that are conducted annually by the Department of Fisheries and Oceans (DFO) of the eastern Scotian Shelf (Supplementary Fig. 1), Canada[31–34]. Results from a secondary, fishery-independent bottom trawl survey beginning in 1986 and conducted during the month of March were used to assess the benthic fish biomass dynamics further. We consider this a secondary survey because of its reduced geographic coverage, different statistical design, and three missing/incomplete years in comparison to the July–August survey.

The dominant, large-bodied predator species, designated the predator group, are Atlantic cod *Gadus morhua* (L_{max} = 161 cm), haddock *Melanogrammus aeglefinus* (87 cm), pollock *Pollachius virens* (112 cm), longfin hake *Urophycis chesteri* (71 cm), silver hake *Merluccius bilinearis* (73 cm), white hake *Urophycis tenuis* (142 cm), red hake *Urophycis chuss* (69 cm), redfish *Sebastes* spp. (60 cm), thorny skate *Amblyraja radiata* (120 cm), spiny dogfish *Squalus acanthias* (196 cm), Greenland halibut *Reinhardtius hippoglossoides* (82 cm), American plaice *Hippoglossoides platessoides* (76 cm), winter flounder *Pseudopleuronectes americanus* (64 cm), witch flounder *Glyptocephalus cynoglossus* (67 cm), and yellowtail flounder *Limanda ferruginea* (59 cm). These fifteen species have been commercially exploited, often in mixed fisheries, throughout the continental shelf in depths of less than 200 m. Ages are determined for Atlantic cod, haddock, pollock and silver hake; a growth model for redfish[35] applicable to the eastern Scotian Shelf stock and an age–length key for American plaice[36] permitted approximations of abundance at age for these two species. The availability of age data permitted the estimation of total mortalities (Supplementary Table 1) and growth rates (Supplementary Table 2).

The forage species group consists of three species: herring *Clupea harengus*, capelin *Mallotus villosus*, and northern sand lance *Ammodytes dubius*, which are under-sampled by the bottom trawl survey. Correction factors have been applied (capelin and sand lance by a factor of 200 and herring by 40 (ref. 37)). Commercial exploitation of these species has been relatively low (herring) or non-existent in this region. Annual anomalies (s.d. units) of individual average body weights (total biomass/total abundance) for each of the three forage species were estimated and used as indices of the temporal variation in physiological condition.

Lower trophic level data (phytoplankton colour index and zooplankton abundance time series) were obtained from Continuous Plankton Recorder (CPR) observations collected on the Scotian Shelf beginning in 1961 at a nominal sampling interval of 1 month. The CPR instrument is towed behind a vessel at about 7 m depth; plankton are collected on a 270 µm silk mesh over a 10 nautical mile tow and stored within the recorder for later identification. Details of the sampling and analysis methods can be found on the Sir Alister Hardy Foundation website (http://www.sahfos.ac.uk/).

The number of CPR samples per month by year is shown in Supplementary Table 3. The series consists of two periods of data collection, 1961–1976, when there were many months without samples, and 1991–2008, when most months were sampled. There was a large data gap from 1977–1990. To compensate for missing monthly data, a time series for the period 1961–1976 was created by averaging monthly values over successive 5-year periods. For the better-sampled 1991–2008 period, 3-year averaging blocks were used. The averaging acts like a rough low-pass filter. The positions of all samples are shown in Supplementary Fig. 9.

Zooplankton data obtained from the CPR Program were grouped into large-bodied (≥2 mm carapace length) and small-bodied (<2 mm) species. The large-bodied category comprised all *Calanus* species and their copepodite stages including *C. finmarchicus*, *C. hyperboreus*, *C. helgolandicus* and *C. glacialis*; all *Metridia*, *Euchaeta* and *Pleuromamma* species including *M. lucens*, *M. longa*, *E. marina*, *E. norvegica*, *P. abdominalis* and *P. robusta*; also included were *Candacia* sp., *C. armata*, *Heterorhabdus* sp. and *H. papilliger*, and *Neocalanus gracilis*, *Rhincalanus nasutus*, *Euchirella rostrata* and *Anomalocera patersoni*.

The small-bodied category included: all *Centropages* and *Temora* species including *C. typicus*, *C. hamatus*, *C. bradyi*, *C. chierchiae*, *T. longicornis* and *T. stylifera*; also *Pseudocalanus*, *Candacia*, *Paracalanus*, *Oncaea*, *Sapphirina*, *Lucicutia*, *Scolecithricella*, *Clausocalanus* and *Calocalanus* sp.; and *Acartia* sp., *Acartia danae*, *Corycaeus* sp., *Ctenocalanus vanus*, *Mecynocera clausi*, *Nannocalanus minor*, *Pleuromamma* sp., *Pleuromamma borealis*, *Lucicutia* and *Tortanus discaudatus*. Further details of the species identification protocols used are provided elsewhere[38].

Bottom trawl survey precision. The precision of the biomass estimates at both the functional group and species level was evaluated by using the random, stratified survey design (27 strata) to calculate the relative standard error (rse, standard error divided by the mean and expressed as a percentage)[39]. For the aggregate predator biomass, the rse (averaged over 41 years) equalled 17.5% (range: 8–40%). The rse for the annual forage fish biomass was averaged over a fewer number of surveys (1993 to 2010) because, before 1993, there were very low levels of forage

fish with the majority of survey sets yielding null values (>70%). The resulting rse was 32% (range: 16–59%). The rse of the annual biomasses of the individual predator species are shown in Supplementary Fig. 6. The average value for cod-like species (35%) compares favourably to surveys in other geographic regions[40]. The forage species biomass estimates averaged 53, 45 and 38%, for capelin, northern sand lance and herring, respectively.

We also considered the probability that the aggregate benthic fish biomass from the relatively constant period of 2000–2004 (168,000 t) was different from the 2010 value (336,000 t) represented by a linear fit to the 2004–2010 observations when the biomass had an average increase of 28,000 t per year. From the analysis above, we took 19% as a representative rse and computed the probability that the 2000–2004 value differed from the 2010 best-fit value. The probability that the 2000–2004 average value was greater than or equal to the best-fit 2010 value was 0.009.

Commercial landings and fishery regulations. Annual eastern Scotian Shelf landings for the benthic species were extracted from databases maintained by the Northwest Atlantic Fisheries Organization (NAFO) and the DFO. The management unit, NAFO Div. 4VW (Supplementary Fig. 1), goes beyond the continental shelf into offshore slope waters where fishing effort was concentrated for certain species, specifically silver hake. Therefore, estimation of meaningful exploitation levels based on the ratio of landings to biomass (derived from the fishery-independent trawl surveys that are constrained to the continental shelf) required the exclusion of the landings data for silver hake. No survey catchability corrections were made for any of the benthic fish species and this may result in annual ratios greater than 1. Hence, this measure of exploitation, often referred to as relative *F* (fishing mortality), is meant to serve as an index of commercial exploitation (Fig. 1c). There are no contemporary (post-2000) estimates of instantaneous fishing mortality rates because of the low levels of fishing associated with the cod and haddock fishing moratorium; for eleven species that were not aged there have never been any estimates of instantaneous fishing mortality rates.

In September of 1993 the eastern Scotian Shelf fishery for cod and haddock was closed while fisheries for flatfish, skates, redfish, silver hake (omitted) and some other minor species have remained open and, since then, total landings have averaged ~6,000 t. For silver hake, landings averaged 16,000 t during this same period and the percentage of cod by-catch associated with this directed fishery was very low (0.01%). The second leading fishery in this area was for redfish (4,500 t) and it incurred by-catches of cod of 0.8%. This information is available at http://www.dfo-mpo.gc.ca/fm-gp/initiatives/cod-morue/strategic-mar-eng.htm.

Principal components analysis and regime shift detection. The status of the eastern Scotian Shelf ecosystem was characterized by principal components analysis based on standardized anomalies (s.d. units) from five biological time series (CPR phytoplankton colour index, zooplankton (body size < and ≥ 2 mm), forage fish and large benthic predators) using a method previously described[41]. PCA1 and 2 accounted for 56.3% and 18% of the variance, respectively. We compared the time series variability of PCA1 constructed from the correlation matrices based on the data series from 1970–2010, which include data gaps for the three lower trophic levels from 1975–1991 and 2008–2009 and from the period (1970–74; 1992–2007) when all series were complete. The r^2 between the two PCA1 series was 0.99. The loadings based on the 1970–2009 (1970–1974; 1992–2008) series of the five variables on PCA1 were: predators = 0.49 (0.50), forage fish complex = −0.44 (−0.46), large-bodied zooplankton = 0.43 (0.43), small-bodied zooplankton = −0.33 (−0.32), phytoplankton colour index = −0.52 (−0.51).

To objectively determine whether and when ecosystem transitions occurred, the dominant principal component axis (PCA1) from the five-variable biological analysis was subjected to a sequential *t*-test analysis of regime shift method (STARS) that identifies the magnitude and direction of shifts significant at a pre-determined α-level, given both the expected cut-off length for the regime (*L*) and a parameter that designates how much of a difference from the observed mean (in s.d. units) are required before data are considered outliers[29] (available at http://www.beringclimate.noaa.gov/regimes/index.html). Choice of *L* and α affect the number and duration of regime shifts[29]. We set *L* = 5 and α = 0.01, in keeping with previous analyses using a subset of ecosystem data where shifts in Scotian Shelf cod–prey states were shown to be relatively insensitive to changes in *L* and *H* (a parameter describing the treatment of outliers) and where the relatively stringent α-value was established to limit regime shift detection to only cases with strong evidence[42]. In our analyses, *L* = 10, *L* = 15 did not affect the timing of the biological regime shifts; setting α = 0.05 only suggested one additional small regime shift in 1975 when *L* = 5.

Carrying capacity estimation. Estimates of the zooplankton biomass on the eastern Scotian Shelf have been compiled[43]; additional stations were available through the BIOCHEM database (http://www.meds-sdmm.dfo-mpo.gc.ca/biochem/Biochem_e.htm) and courtesy of C. Johnson and B. Casault. Biomass estimates were made using vertical hauls from the bottom to the surface using 200 µm mesh nets. Observations from 1999–2008 provided reasonable spatial

coverage for the months of March, April, July and October. March and July observations were made during ecosystem fisheries surveys; April and October data, from Atlantic Zone Monitoring Program cruises, were collected on three standard, repeated sections at the northeastern (Cabot Strait Section), central (Louisbourg Section) and southwestern end (Halifax Section) of NAFO Div. 4VW (Supplementary Fig. 10).

The zooplankton biomasses are shown in Supplementary Table 4. Taking the simple average of these observations and multiplying by the surface area of the eastern shelf gives a zooplankton standing crop of 4.8×10^6 t. Previous work[44] estimated a zooplankton production to biomass ratio of 6 to 9.7, and used an average value of 7 as an overall factor applied to zooplankton as a group. Another more rigorous, quantitative analysis[45], estimated production to biomass (P/B) ratios for nine species, including leading components of the overall biomass, and for a single general category. Their biomass-weighted P/B ratio was 8.9. Multiplying the average biomass in Supplementary Table 4 by this factor yields an overall zooplankton production of approximately 4.3×10^7 t. Using a 10% efficiency relating the zooplankton production to forage fish yields a rough estimate of the forage fish carrying capacity of 4.3×10^6 t.

Similar to our calculation of carrying capacity for forage species, we estimated the annual production of phytoplankton on the eastern Scotian Shelf based on measurements from three sources[46–48]. The data from two sources[46,47] covered the months of March–August and November; the observations from the third[47] consisted of 13 monthly estimates from March 1991 to March 1992. All observations were reported as mg of carbon per m^2 per hour. We converted mg C to wet weight of phytoplankton using a factor of 42 (ref. 49), thus allowing direct comparison with our estimates of zooplankton production. Combining two sets[46,47] of results, we estimate the annual production of phytoplankton as 640×10^6 t and 240×10^6 t, respectively. These give ratios of phytoplankton to zooplankton production of 15 and 5.5 which are reasonable if the energy transfer efficiency is ~10%. One crude measure for estimating the carrying capacity for the predator complex involved taking the ratio of peak biomasses of benthic fish predators to forage fishes which is about 1:16.

Damped harmonic oscillator calculations. The input data set was the biomass of the three leading forage species. The period under consideration was 1994–2010. We used a temporally varying background state derived by running a 30% linear LOESS filter through the data. The filter split the variability into two components: a very low frequency component with a period estimated as 56 years from the autocorrelation function and higher frequency variability (Supplementary Fig. 8).

The anomalies indicated a simple damped harmonic oscillator-like (dSHO) variability, particularly from 1994 to 2001, shown as solid grey circles in Fig. 3. We solved for the characteristics of this variability by least squares fitting the observations to the dSHO equation:

$$A_0 + A_1^{(-d_1 t)} \cos(2\pi t/\tau),$$

where A_0 is the mean, A_1 the amplitude of the harmonic oscillation, d_1 the dampening rate, and τ the periodicity of the oscillation. The solution based on optimal fitting of 1994–2001 observations, shown as a solid red line in Fig. 3, gives $A_0 = 347{,}000$ t, $A_1 = 6{,}100{,}000$ t, $1/d_1 = 6.9$ years, $\tau = 5.2$ years.

Environmental indices and their relationship to the pre-collapse, collapse and post-collapse periods. The relationship between forage fish biomass and water column temperatures was evaluated. The results showed no relationship between the two variables at 0 lag (Supplementary Fig. 5). We also lagged the forage fish biomass relative to the temperature; for lags up to 4 years, the r^2 was less than 0.04. At lags of 5 years, $r^2 = 0.21$; however, this is approaching the life cycle of these species and the biological import of the enhanced correlation at this lag is therefore questionable. Finally, we examined integrated temperatures up to 5 years. All r^2 values were less than 0.02.

An examination of the seasonal variability of the stratification index showed that the long-term trend seen in the annual data was due mostly to the summer and secondarily to the fall series (Supplementary Fig. 11). Moreover, the magnitude of the changes during winter and spring (major bloom period) were small compared to those in summer and fall. The period, 1985–1994, of ongoing decline of the benthic fish functional group showed near-normal values of the stratification parameter during the spring bloom period. The outbreak period (1994–2001) of the forage group was weakly ($r^2 = 0.21$) related (not significantly) to the spring stratification.

The seasonal variation of the phytoplankton colour index (Supplementary Fig. 12) and their correlations with the annual index, their averages and deviations

(Supplementary Table 5) indicated that the greatest contributions were from the winter and spring seasons (high correlations and variance), followed by fall (high correlation, moderate variance), and finally by summer (low correlation, lowest variance). This indicated that summer, with the greatest stratification changes, had the least impact on the annual colour index; fall, with the next greatest impact on stratification, made the second least contribution to the index. Further comparisons of the monthly colour indexes with the annual revealed that the greatest contributions were from March and April, typical months for the spring bloom on the eastern Scotian Shelf (Supplementary Table 6).

We assessed the impact of the full range of temperatures observed in the eastern Scotian Shelf using the relationship between the intrinsic rate of population growth, r, age at maturity, a, and bottom temperatures developed previously for 20 stocks of North Atlantic cod[50]. This analysis revealed that the estimates of r and a so derived were, during the pre-collapse period, on average +4% higher and −3% lower, respectively, than during the collapsed period and +3% higher and −2% lower, respectively, for the recovering period. We also note that the period 1987–1991, immediately preceding the cod collapse, featured some of the coldest bottom water temperatures on record. Again applying the approach detailed above, and the r and a data provided[50], we estimate a maximum 10% decrease in r and a 9% increase in a relative to the collapsed and post-collapse intervals. During this cold period, fishing mortality on cod, based on the ratio of landings to biomass, averaged 65%, which is an extremely high annual rate of biomass removal. This dwarfs any possible impact of the environment as expressed through the effect of temperature on r and a.

The population dynamics and feeding ecology of grey seals on the eastern Scotian Shelf has been assessed since the early 1960s with sampling effort concentrated on the colony inhabiting Sable Island. Total population sizes were estimated from a model fit to census data on pup production (Supplementary Fig. 13).

31. Doubleday, W. G. *Studies No. 2, Manual on Groundfish Surveys in the Northwest Atlantic* (NAFO Scientific Council Studies, 1981).
32. Doubleday, W. G. & Rivard, D. Bottom trawl surveys. *Can. Spec. Pub. Fish. Aquat. Sci.* **58,** 1–273 (1981).
33. Chadwick, E. M. P. *et al.* History of annual multi-species trawl surveys on the Atlantic coast of Canada. *AZMP Bull.* **6,** 25–42 (2007).
34. Shackell, N. L. & Frank, K. T. Marine fish diversity on the Scotian Shelf, Canada. *Aquat. Conserv.: Mar. Freshwat. Ecosyst.* **13,** 305–321 (2003).
35. Mayo, R. K., Gifford, V. M. & Jearld, A. Age validation of redfish, *Sebastes marinus* (L.), from the Gulf of Maine-Georges Bank Region. *J. Northwest Atl. Fish. Sci.* **2,** 13–19 (1981).
36. Fowler, G. W. & Stobo, W. T. Status of 4VW American plaice and yellowtail flounder. *DFO Can. Sci. Advis. Sec. Res. Doc.* **2000,** 1–86 (2000).
37. Bundy, A. Mass balance models of the eastern Scotian Shelf before and after the cod collapse and other ecosystem changes. *Can. Tech. Rep. Fish. Aquat. Sci.* **2520,** 1–192 (2004).
38. Richardson, A. J. *et al.* Using continuous plankton recorder data. *Prog. Oceanogr.* **68,** 27–74 (2006).
39. Smith, S. J. Analysis of data from bottom trawl surveys. *NAFO Sci. Coun. Studies* **28,** 25–53 (1996).
40. Benoît, H., Swain, D. & Chouinard, G. Using the long-term bottom-trawl survey of the Southern Gulf of St. Lawrence to understand marine fish populations and community change. *AZMP Bull.* **8,** 19–27 (2008).
41. Choi, J. S., Frank, K. T., Petrie, B. D. & Leggett, W. C. Integrated assessment of a large marine ecosystem: a case study of the devolution of the Eastern Scotian Shelf, Canada. *Oceanogr. Mar. Biol. Ann. Rev* **42,** 47–67 (2005).
42. Litzow, M. A., Urban, J. D. & Laurel, B. J. Increased spatial variance accompanies reorganization of two continental shelf ecosystems. *Ecol. Apps* **18,** 1331–1337 (2008).
43. Harrison, G. *et al.* Optical, chemical, and biological oceanographic conditions in the Maritimes Region in 2008. *DFO Can. Sci. Advis. Sec. Res. Doc.* **2009,** 1–55 (2009).
44. Mills, E. L. & Fournier, R. O. Fish production and the marine ecosystems of the Scotian Shelf, eastern Canada. *Mar. Biol.* **54,** 101–108 (1979).
45. Tremblay, M. J. & Roff, J. C. Production estimates for Scotian Shelf copepods based on mass specific P/B ratios. *Can. J. Fish. Aquat. Sci.* **40,** 749–753 (1983).
46. Cochlan, W. Seasonal study of uptake and regeneration of nitrogen on the Scotian Shelf. *Cont. Shelf Res.* **5,** 555–577 (1986).
47. Fournier, R., Marra, J., Bohrer, R. & van Det, M. Plankton dynamics and nutrient enrichment of the Scotian Shelf. *J. Fish. Res. Bd Can.* **34,** 1004–1018 (1977).
48. Mousseau, L., Fortier, L. & Legendre, L. Annual production of fish larvae and their prey in relation to size-fractionated primary production (Scotian Shelf, NW Atlantic). *ICES J. Mar. Sci.* **55,** 44–57 (1998).
49. Cauffopé, G. & Heymans, S. Energy contents and conversion factors for sea lion's prey. *UBC Fish. Centre Res. Rep* **13,** 225–237 (2005).
50. Myers, R. A., Mertz, G. & Fowlow, P. S. Maximum population growth rates and recovery times for Atlantic cod, *Gadus morhua*. *Fish. Bull.* **95,** 762–772 (1997).

LETTER

The ageing systemic milieu negatively regulates neurogenesis and cognitive function

Saul A. Villeda[1,2], Jian Luo[1], Kira I. Mosher[1,2], Bende Zou[3], Markus Britschgi[1]†, Gregor Bieri[1,4], Trisha M. Stan[1,5], Nina Fainberg[1], Zhaoqing Ding[1,5], Alexander Eggel[1], Kurt M. Lucin[1], Eva Czirr[1], Jeong-Soo Park[1]†, Sebastien Couillard-Després[6], Ludwig Aigner[6], Ge Li[7], Elaine R. Peskind[7,8], Jeffrey A. Kaye[9], Joseph F. Quinn[9], Douglas R. Galasko[10], Xinmin S. Xie[3], Thomas A. Rando[1,11,12] & Tony Wyss-Coray[1,2,5,11]

In the central nervous system, ageing results in a precipitous decline in adult neural stem/progenitor cells and neurogenesis, with concomitant impairments in cognitive functions[1]. Interestingly, such impairments can be ameliorated through systemic perturbations such as exercise[1]. Here, using heterochronic parabiosis we show that blood-borne factors present in the systemic milieu can inhibit or promote adult neurogenesis in an age-dependent fashion in mice. Accordingly, exposing a young mouse to an old systemic environment or to plasma from old mice decreased synaptic plasticity, and impaired contextual fear conditioning and spatial learning and memory. We identify chemokines—including CCL11 (also known as eotaxin)—the plasma levels of which correlate with reduced neurogenesis in heterochronic parabionts and aged mice, and the levels of which are increased in the plasma and cerebrospinal fluid of healthy ageing humans. Lastly, increasing peripheral CCL11 chemokine levels *in vivo* in young mice decreased adult neurogenesis and impaired learning and memory. Together our data indicate that the decline in neurogenesis and cognitive impairments observed during ageing can be in part attributed to changes in blood-borne factors.

Adult neurogenesis occurs in local microenvironments, or neurogenic niches, in the subventricular zone (SVZ) of the lateral ventricles and the subgranular zone (SGZ) of the hippocampus[2,3]. Permissive cues within the neurogenic niche are thought to drive the production of new neurons and their subsequent integration into the neurocircuitry of the brain[4,5], directly contributing to cognitive processes including learning and memory[6–9]. Importantly, the neurogenic niche is localized around blood vessels[10,11], allowing for potential communication with the systemic environment. Therefore, the possibility arises that diminished neurogenesis during ageing may be modulated by the balance of two independent forces: intrinsic central nervous system (CNS)-derived cues[12–14] and cues extrinsic to the CNS delivered by blood. Thus we hypothesized that age-related systemic molecular changes could cause a decline in neurogenesis and impair cognitive function during ageing.

We first characterized cellular, electrophysiological and behavioural changes associated with the neurogenic niche in the dentate gyrus of the hippocampus in an ageing cohort of mice. We observed cellular changes consistent with markedly decreased adult neurogenesis[1] and increased neuroinflammation with age[15] (Supplementary Fig. 2a–e). Additionally, we detected deficits in synaptic plasticity (Supplementary Fig. 3a–c), and behavioural deficits in contextual fear conditioning (Supplementary Fig. 4a–c) and radial arm water maze (RAWM; Supplementary Fig. 4d–f) paradigms in old animals, consistent with decreased cognitive function during ageing[16].

Next we investigated the contribution of peripheral systemic factors to the age-related decline in neurogenesis in the dentate gyrus of the hippocampus in the setting of isochronic (young–young and old–old) and heterochronic (young–old) parabiosis (Fig. 1a). Remarkably, the number of doublecortin (Dcx)-positive newly born neurons (Fig. 1b, c), BrdU-positive cells (Fig. 1e, f) and Sox2-positive progenitors (Supplementary Fig. 5a, b) decreased in young heterochronic parabionts. In contrast, we observed an increase in the number of Dcx-positive (Fig. 1b, d), BrdU-positive (Fig. 1e, g) and Sox2-positive (Supplementary Fig. 5a, c) cells in the old heterochronic parabionts. The number of Dcx-positive neurons between unpaired age-matched animals and isochronic animals showed no difference (Supplementary Fig. 6a, b). As a control, flow cytometry analysis confirmed a shared vasculature in a subset of parabiotic pairs, in which one parabiont was transgenic for green fluorescent protein (GFP; Supplementary Fig. 7a–d). Together our findings suggest that global age-dependent systemic changes can modulate neurogenesis in both the young and aged neurogenic niche, potentially contributing to the decline in regenerative capacity observed in the normal ageing brain.

As previously reported by others[17], we rarely detected peripherally derived GFP cells in the CNS of wild-type mice, and these numbers did not differ between isochronic and heterochronic pairings (Supplementary Fig. 7e), suggesting that the observed effects are mediated by soluble factors in plasma. To confirm that circulating factors within aged blood contribute to reduced neurogenesis with age, we intravenously injected plasma isolated from young or old mice into young adult mice (Fig. 2a). The number of Dcx-positive cells in the dentate gyrus decreased in animals receiving old plasma compared to animals receiving young plasma (Fig. 2b, c), indicating that soluble factors present in old blood inhibit adult neurogenesis.

To investigate the functional effect of the ageing systemic milieu, extracellular electrophysiological recordings were done on hippocampal slices prepared from young isochronic and heterochronic parabionts (Fig. 1h, i and Supplementary Fig. 6c). A decrease in long-term potentiation (LTP) in the dentate gyrus of heterochronic parabionts was detected (Fig. 1h, i). These data indicate that age-related systemic changes can elicit deficits in synaptic plasticity. As LTP is considered to be a correlate of learning and memory[18], these finding suggest that age-related systemic changes may also affect cognitive functions during ageing.

[1]Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, California 94305, USA. [2]Neuroscience IDP Program, Stanford University School of Medicine, Stanford, California 94305, USA. [3]AfaSci Research Laboratory, Redwood City, California, 94063, USA. [4]School of Life Sciences, Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland. [5]Immunology IDP Program, Stanford University School of Medicine, Stanford, California 94305, USA. [6]Institute of Molecular Regenerative Medicine, Paracelsus Medical University, Strubergasse 21, A-5020 Salzburg, Austria. [7]Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, Washington 98108-1597, USA. [8]Veterans Affairs Northwest Network Mental Illness Research, Education, and Clinical Center, Seattle, Washington 98108-1597, USA. [9]Layton Aging & Alzheimer's Disease Center, Oregon Health and Science University, CR131, 3181 SW Sam Jackson Park Road, Portland, Oregon 97201-3098, USA; and Portland VA Medical Center, Portland, Oregon 97207, USA. [10]Department of Neurosciences, University of California San Diego, 9500 Gilman Drive #0948, La Jolla, California 92093-0948, USA. [11]Center for Tissue Regeneration, Repair and Restoration, VA Palo Alto Health Care System, Palo Alto, California 94304, USA. [12]The Glenn Laboratories for the Biology of Aging, Stanford University School of Medicine, Stanford, California 94305, USA. †Present addresses: CNS Discovery, pRED, F. Hoffmann-La Roche Ltd., CH-4070 Basel, Switzerland (M.B.); Department of Biochemistry, College of Medicien, Dankook University, Cheonan 330-714, South Korea (J.S.P.).
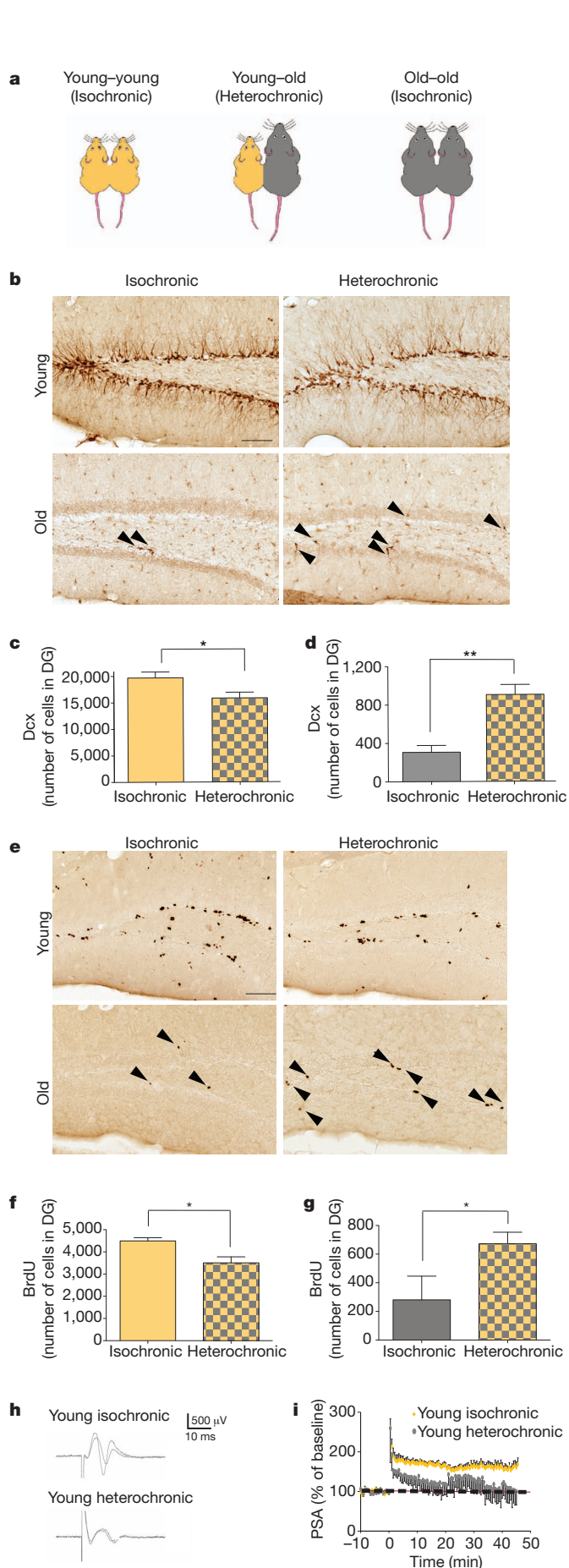
Figure 1 | Heterochronic parabiosis alters neurogenesis in an age-dependent fashion. a, Schematic showing parabiotic pairings. **b**, **e**, Representative fields of Dcx (**b**) and BrdU (**e**) immunostaining of young (3–4 months; yellow) and old (18–20 months; grey) isochronic and heterochronic parabionts 5 weeks after parabiosis (arrowheads point to individual cells; scale bars, 100 μm). **c**–**f**, Quantification of neurogenesis (**c**, **d**) and proliferating cells (**e**, **f**) in the young (**c**, **e**; top) and old (**d**, **f**; bottom) dentate gyrus (DG) after parabiosis. Data from 12 young isochronic, 10 young heterochronic, 6 old isochronic and 12 old heterochronic parabionts. **g**, **h**, Population spike amplitude (PSA) was recorded from the dentate gyrus of young parabionts. Representative electrophysiological profiles (**g**) and LTP levels (**h**) are shown for young heterochronic and isochronic parabionts. Data from 4–5 mice per group. All are data represented as mean + s.e.m.; *$P < 0.05$; **$P < 0.01$, t-test.
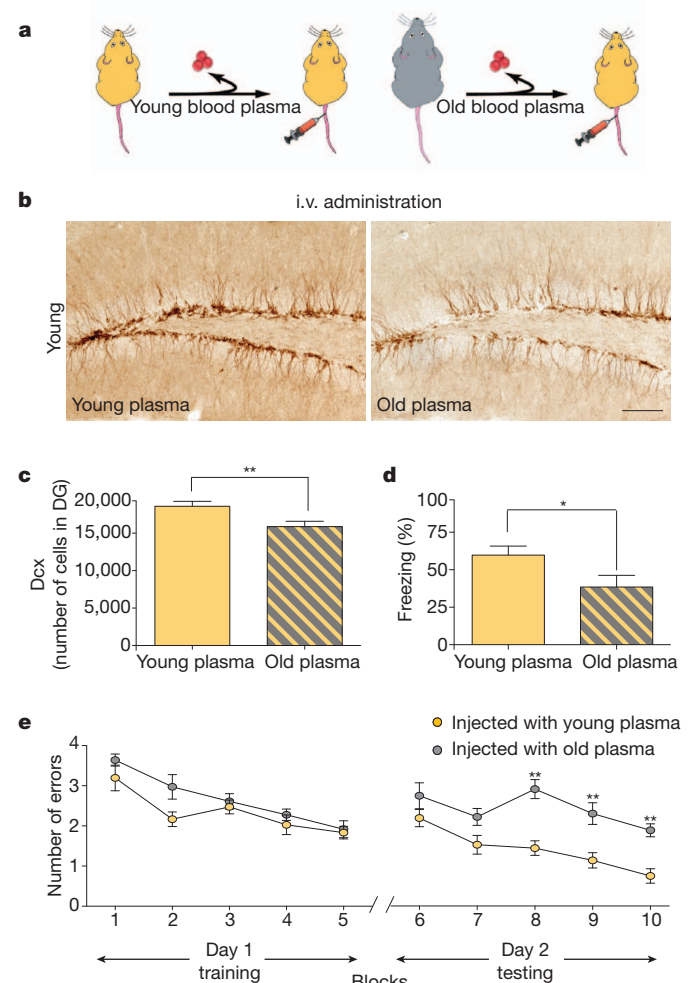
Figure 2 | Factors from an old systemic environment decrease neurogenesis and impair learning and memory. a, Schematic of young (3–4 months) or old (18–22 months) plasma extraction and intravenous (i.v.) injection into young (3 months) adult mice. **b**, Representative field of Dcx immunostaining of young adult mice after plasma injection treatment four times over 10 days (scale bar, 100 μm). **c**, Quantification of neurogenesis in the young dentate gyrus after plasma injection. Data from 8 mice injected with young plasma and 7 mice injected with old plasma. **d**, **e**, Hippocampal learning and memory assessed by contextual fear conditioning (**d**) and RAWM (**e**) paradigms in young adult mice after young or old plasma injections nine times over 24 days. **d**, Percent freezing time 24 h after training. Data from 8 mice per group. **e**, Number of entry arm errors before finding platform. Data from 12 mice per group. All data represented as mean ± s.e.m.; *$P < 0.05$; **$P < 0.01$, t-test (**c**, **d**), repeated measures ANOVA, Bonferroni post-hoc test (**e**).

1 SEPTEMBER 2011 | VOL 477 | NATURE | 91

Subsequently, we tested hippocampal-dependent learning and memory using contextual fear conditioning and RAWM paradigms in young adult mice intravenously injected with young or old plasma (Fig. 2d, e). During fear conditioning training mice exhibited no differences in baseline freezing regardless of plasma injection treatment (Supplementary Fig. 8a). However, mice receiving old plasma demonstrated decreased freezing in contextual (Fig. 2d), but not cued (Supplementary Fig. 8b), memory testing. During the training phase of the RAWM all mice showed similar swim speeds (Supplementary Fig. 8c) and spatial learning capacity for the task (Fig. 2e). However, during the testing phase animals that had received old plasma demonstrated impaired learning and memory for platform location (Fig. 2e). As a control, we tested the RAWM paradigm in young adult mice with ablated hippocampal neurogenesis and observed corresponding behavioural deficits (Supplementary Fig. 9a–e). Collectively, these data indicate that factors present in ageing blood inhibit adult neurogenesis, and functionally contribute to impairments in cognition.

Previous pioneering studies focused on muscle have shown that exposure of the aged stem cell niche to a young systemic environment through heterochronic parabiosis results in increased regeneration after muscle injury[19], in part by involving Notch signalling[28]. However, individual systemic factors associated with ageing and tissue degeneration have not yet been characterized or investigated for their role in regulating the decline in tissue regeneration. To identify such factors, we used a proteomic approach in which relative levels of 66 cytokines, chemokines and other secreted signalling proteins were measured in the plasma of normal ageing mice using standardized multiplex sandwich enzyme-linked immunosorbent assays (ELISAs; Luminex) (Supplementary Table 1). Using multivariate analysis, we identified seventeen proteins whose levels increased and correlated with decreased neurogenesis during ageing (Fig. 3a and Supplementary Fig. 10a, b). To identify systemic factors associated with heterochronic parabiosis, we analysed plasma samples from young and old mice before and after pairings in an independent proteomic screen. Comparison of young isochronic and heterochronic cohorts identified fifteen factors that increased in heterochronic parabionts (Fig. 3a and Supplementary Fig. 10c), whereas comparison between old isochronic and heterochronic cohorts revealed four factors that decreased in isochronic parabionts (Supplementary Fig. 10c). Interestingly, only six factors—CCL2, CCL11, CCL12, CCL19, haptoglobin and β2-microglobulin—were elevated in old unpaired and young heterochronic cohorts (Fig. 3a). Of these, CCL11 is a chemokine involved in allergic responses and not previously linked to ageing, neurogenesis or cognition. Relative levels of CCL11 were increased in the plasma of mice during normal ageing (Fig. 3b) and in young mice during heterochronic parabiosis (Fig. 3c). Furthermore, we detected an age-related increase in CCL11 in plasma (Fig. 3d) and cerebrospinal fluid (CSF; Fig. 3e), from healthy human individuals between 20 and 90 years of age, suggesting that this age-related systemic increase is conserved across species.

Having identified CCL11 as an age-related systemic factor associated with decreased neurogenesis, we tested its potential biological relevance *in vivo*. We administered CCL11 protein through intraperitoneal injections into young adult Dcx-luciferase reporter mice[20], and using a non-invasive bioluminescent imaging assay detected a significant decrease in neurogenesis (Supplementary Fig. 11b, c). Using immunohistochemical analysis we investigated the effect of systemic CCL11 on adult hippocampal neurogenesis in young adult wild-type mice. We administered CCL11 or vehicle alone, and in combination with either an anti-CCL11 neutralizing antibody or an isotype control antibody through intraperitoneal injections (Fig. 4a). Systemic administration of CCL11 induced an increase in CCL11 plasma levels (Supplementary Fig. 11a), and significantly decreased the number of Dcx-positive cells in the dentate gyrus (Fig. 4b, c). Importantly, this decrease in neurogenesis could be rescued by systemic neutralization of CCL11 (Fig. 4b, c). Likewise, BrdU-positive cells also showed similar changes in cell number (Supplementary Fig. 11d, e), and furthermore
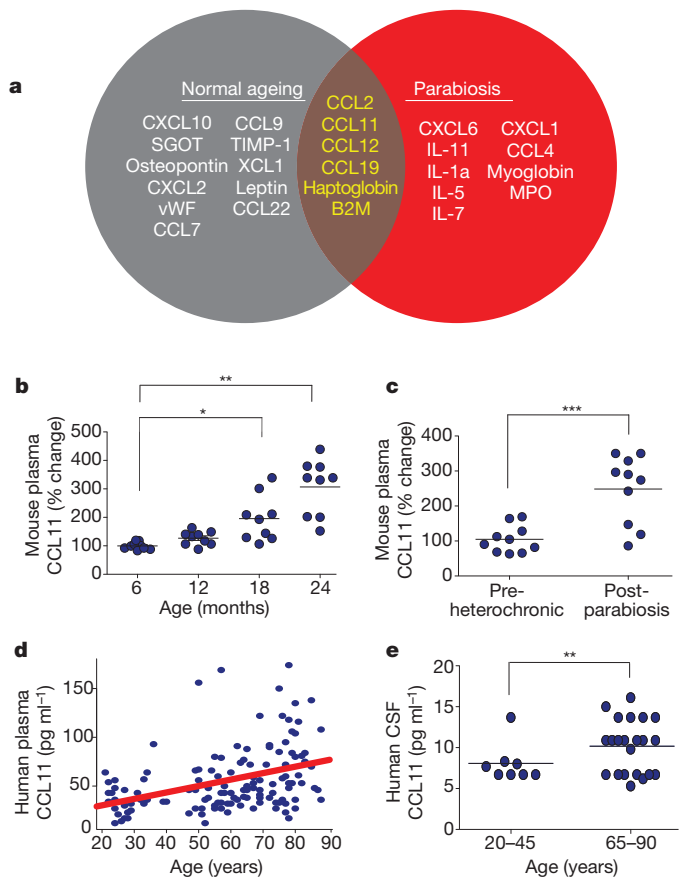
Figure 3 | Systemic chemokine levels increase during ageing and heterochronic parabiosis, and correlate with decreased neurogenesis. a, Venn diagram of results from ageing and parabiosis proteomic screens. In grey are shown the seventeen age-related plasma factors that correlated most strongly with decreased neurogenesis, in red are shown the fifteen plasma factors that increased between young isochronic and young heterochronic parabionts, and in the brown intersection are the six factors elevated in both screens. Data from 5–6 mice per age group. **b, c**, Changes in plasma concentrations of CCL11 with age (**b**) and young heterochronic parabionts pre- and post- parabiotic pairing (**c**). **d, e**, Changes in plasma (**d**; $r = 0.40$; $P = 5.6 \times 10^{-7}$; 95% confidence interval = 0.26–0.53) and CSF (**e**) concentrations of CCL11 with age in healthy human subjects. All data represented as dot plots with mean; $*P < 0.05$; $**P < 0.01$; $***P < 0.001$, t-test (**c, e**), ANOVA, Tukey's post-hoc test (**a, b**), and Mann–Whitney U Test (**d**).

the percentage of cells expressing both BrdU and NeuN decreased (Supplementary Fig. 11f, g). The percentage of cells expressing BrdU and GFAP did not significantly change (Supplementary Fig. 11f, h). As a negative control we assayed neurogenesis after systemic administration of monocyte colony stimulating factor (MCSF), a measured protein whose plasma levels do not change with age, and detected no change in Dcx-positive cells in the dentate gyrus (Supplementary Fig. 12a–d). Together, these data indicate that increasing the systemic level of the age-related factor CCL11 is sufficient to partially recapitulate some of the inhibitory effects observed with ageing and heterochronic parabiosis.

Additionally, we investigated the possibility that age-related blood-borne factors influence neural progenitor activity and neural differentiation *in vitro*. We assayed the number of neurospheres formed after exposure of primary neural stem/progenitor cells (NPCs) to serum from aged mice and observed a 50% decrease when compared to exposure to serum from young mice (Supplementary Fig. 13a). We then tested the effect of CCL11 and observed that the number and size of neurospheres formed from primary NPCs exposed to CCL11 significantly decreased (Supplementary Fig. 13b–d). Using a human-derived NTERA cell line expressing enhanced (e)GFP under the Dcx
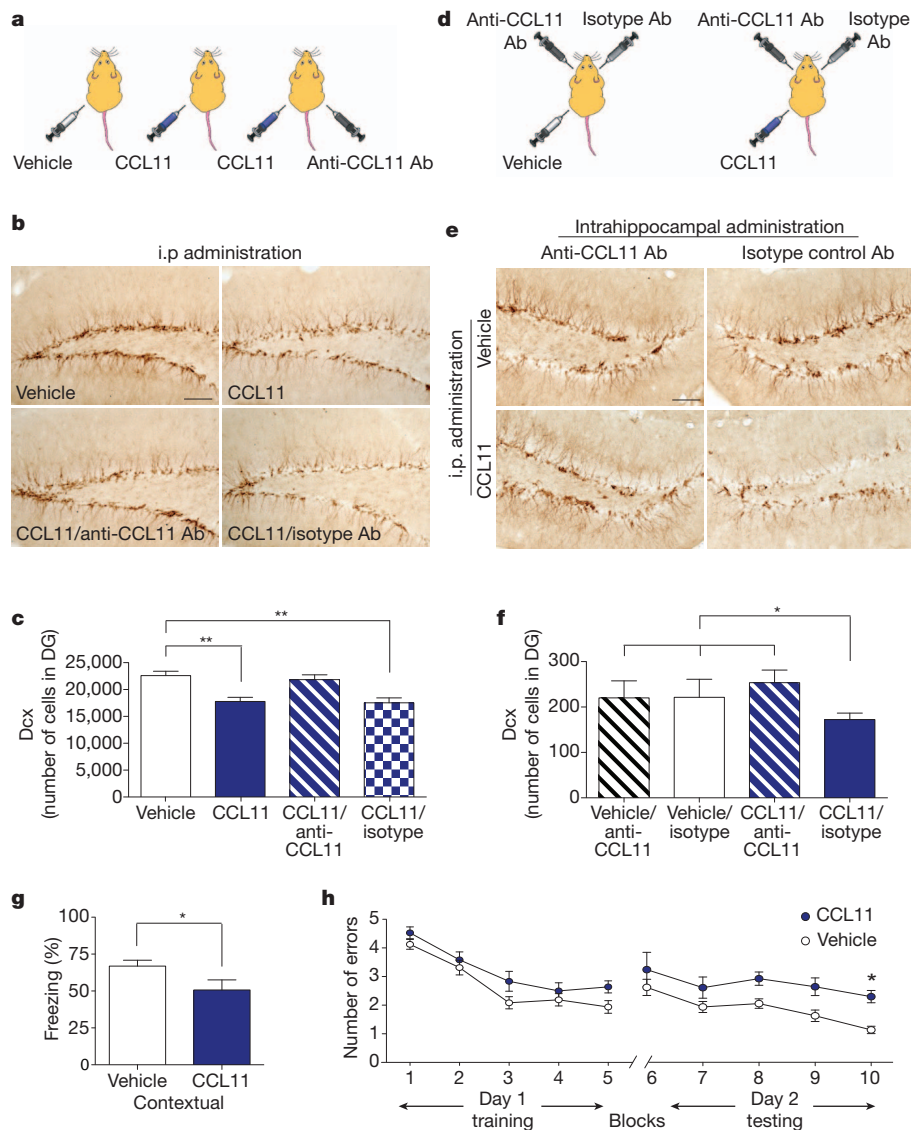
Figure 4 | Systemic exposure to CCL11 inhibits neurogenesis and impairs learning and memory. a, Schematic of young (3–4 months) mice injected intraperitoneally with CCL11 or vehicle, and in combination with anti-CCL11 neutralizing or isotype control antibody (Ab). **b**, Representative field of Dcx-positive cells for each treatment group ($n = 6$–10 mice) treated four times over 10 days. i.p., intraperitoneal. Scale bar, 100 μm. **c**, Quantification of neurogenesis in the dentate gyrus after treatment. **d**, Schematic of young adult mice given unilateral stereotaxic injections of anti-CCL11 neutralizing or isotype control antibody followed by systemic injections with either recombinant CCL11 or PBS (vehicle). **e**, Representative field of Dcx-positive cells in adjacent sides of the dentate gyrus for each treatment group ($n = 3$–11 mice). Scale bar, 100 μm. **f**, Quantification of neurogenesis in the dentate gyrus after systemic and stereotaxic treatment. Bars represent mean number of cells in each section. **g, h**, Learning and memory assessed by contextual fear conditioning (**g**) and RAWM (**h**) paradigms in young adult mice injected with CCL11 or vehicle every 3 days for 5 weeks ($n = 12$–16 mice per group). All data are represented as mean ± s.e.m.; *$P < 0.05$; **$P < 0.01$; ANOVA, Dunnet's or Tukey's post-hoc test (**c**, **f**); repeated measures ANOVA, Bonferroni post-hoc test (**k**).

promoter, we assayed neural differentiation and observed a significant decrease in eGFP expression after 12 days in culture with CCL11 (Supplementary Fig. 13e, f). Although these findings open the possibility of a direct interaction of systemic factors with progenitor cells *in vivo* during ageing, they do not preclude the possibility of indirect actions by interactions with other neurogenic niche cell types.

To examine the effect of CCL11 on neurogenesis in the brain we stereotaxically injected CCL11 into the dentate gyrus of young adult mice, and observed a decrease in the number of Dcx-positive cells compared with the contralateral dentate gyrus, which received vehicle control (Supplementary Fig. 14a, b). Furthermore, we examined whether the inhibitory effect of peripheral CCL11 could be restored locally within the hippocampus. We stereotaxically injected CCL11-specific neutralizing antibody and isotype control antibodies into the contralateral dentate gyrus of young adult mice (Fig. 4d). After stereotaxic injection, we systemically administered CCL11 or vehicle control

by intraperitoneal injections (Fig. 4d). The decrease in Dcx-positive cell number observed in mice that received systemic CCL11 could be rescued by neutralizing CCL11 within the dentate gyrus (Fig. 4e, f), suggesting that the increase in systemic chemokine levels exerts a direct effect in the CNS.

Lastly, to determine the physiological relevance of increased systemic CCL11 levels in mice we assessed hippocampal-dependent learning and memory using contextual fear conditioning and RAWM paradigms (Fig. 4g, h). Young adult mice received intraperitoneal injections of recombinant CCL11 or vehicle control. During fear conditioning training all mice, regardless of treatment, exhibited no differences in baseline freezing (Supplementary Fig. 15a). In contrast, mice that had received CCL11 demonstrated decreased freezing during contextual (Fig. 4g), but not cued (Supplementary Fig. 15b), memory testing. During the training phase of the RAWM task all mice regardless of treatment showed similar swim speeds (Supplementary Fig. 15c) and

learning capacity for the task (Fig. 4h). However, by the end of the testing phase animals that had received CCL11 exhibited impaired learning and memory deficits (Fig. 4h). Together, these functional data demonstrate that increasing the systemic level of CCL11 can not only inhibit adult neurogenesis, but also impair learning and memory.

Cumulatively, our data link age-related molecular changes in the systemic milieu to the age-related decline in adult neurogenesis, and impairments in synaptic plasticity and cognitive function observed during ageing (Supplementary Fig. 1). Whereas local immune signalling in the brain is emerging as a critical modulator of NPC function[11,15,21,22] and neurodegeneration[11,15,21], we now identify systemic immune-related factors as potentially critical contributors to the susceptibility of the ageing brain to cognitive impairments. Interestingly, members of the identified age-related chemokines (CCL2, CCL11 and CCL12) are localized to within 70 kb on mouse chromosome 11, and within 40 kb on human chromosome 17, implicating this genetic locus in normal brain ageing and possibly ageing in general. Indeed, work investigating cellular senescence, a known hallmark of ageing, further suggests the involvement of some of the individual systemic chemokines reported here (CCL2) in the ageing process as components of the senescence-associated secretory phenotype[23]. Lastly, although the proteomic platform we used here was sufficient to identify systemic inhibitory 'ageing' factors it will be critical to develop and utilize broader proteomic screens to facilitate the discovery of systemic pro-neurogenic 'rejuvenating' factors with the potential to ameliorate age-related cognitive dysfunction.

METHODS SUMMARY

Mouse strains used were C57BL/6 (Jackson Laboratory), C57BL/6 aged mice (National Institutes of Ageing), Dcx-Luc[20] and C57BL/6J-Act-GFP (Jackson Laboratory). All animal use was in accordance with institutional guidelines approved by the Veterans Affairs Palo Alto Committee on Animal Research. Parabiosis surgery followed previously described procedures[19] with the addition that peritonea between animals were surgically connected. Immunohistochemistry followed standard published techniques[24]. Extracellular electrophysiology was performed as previously described[25]. Spatial learning and memory was assayed with the RAWM paradigm as previously published[26]. Contextual fear conditioning was assayed as previously published[27]. Relative plasma concentrations of cytokines and signalling molecules in mice and humans were measured using antibody-based multiplex immunoassays at Rules Based Medicine. Statistical analysis was performed with Prism 5.0 software (GraphPad Software). Plasma protein correlations were analysed with the Significance Analysis of Microarray software (*SAM 3.00* algorithm; http://www.stat.stanford.edu/~tibs/SAM/index.htm). Experiments were carried out by investigators blinded to the treatment of animals.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 September 2009; accepted 5 July 2011.

1. van Praag, H., Shubert, T., Zhao, C. & Gage, F. H. Exercise enhances learning and hippocampal neurogenesis in aged mice. *J. Neurosci.* **25,** 8680–8685 (2005).
2. Gage, F. H. Mammalian neural stem cells. *Science* **287,** 1433–1438 (2000).
3. Alvarez-Buylla, A. & Lim, D. A. For the long run: maintaining germinal niches in the adult brain. *Neuron* **41,** 683–686 (2004).
4. Zhao, C., Deng, W. & Gage, F. H. Mechanisms and functional implications of adult neurogenesis. *Cell* **132,** 645–660 (2008).
5. van Praag, H. *et al.* Functional neurogenesis in the adult hippocampus. *Nature* **415,** 1030–1034 (2002).
6. Deng, W., Aimone, J. B. & Gage, F. H. New neurons and new memories: how does adult hippocampal neurogenesis affect learning and memory? *Nature Rev. Neurosci.* **11,** 339–350 (2010).
7. Clelland, C. D. *et al.* A functional role for adult hippocampal neurogenesis in spatial pattern separation. *Science* **325,** 210–213 (2009).
8. Zhang, C. L., Zou, Y., He, W., Gage, F. H. & Evans, R. M. A role for adult TLX-positive neural stem cells in learning and behaviour. *Nature* **451,** 1004–1007 (2008).
9. Saxe, M. D. *et al.* Ablation of hippocampal neurogenesis impairs contextual fear conditioning and synaptic plasticity in the dentate gyrus. *Proc. Natl Acad. Sci. USA* **103,** 17501–17506 (2006).
10. Shen, Q. *et al.* Endothelial cells stimulate self-renewal and expand neurogenesis of neural stem cells. *Science* **304,** 1338–1340 (2004).
11. Carpentier, P. A. & Palmer, T. D. Immune influence on adult neural stem cell regulation and function. *Neuron* **64,** 79–92 (2009).
12. Renault, V. M. *et al.* FoxO3 regulates neural stem cell homeostasis. *Cell Stem Cell* **5,** 527–539 (2009).
13. Molofsky, A. V. *et al.* Increasing $p16^{INK4a}$ expression decreases forebrain progenitors and neurogenesis during ageing. *Nature* **443,** 448–452 (2006).
14. Lie, D. C. *et al.* Wnt signalling regulates adult hippocampal neurogenesis. *Nature* **437,** 1370–1375 (2005).
15. Lucin, K. M. & Wyss-Coray, T. Immune activation in brain aging and neurodegeneration: too much or too little? *Neuron* **64,** 110–122 (2009).
16. Rapp, P. R. & Heindel, W. C. Memory systems in normal and pathological aging. *Curr. Opin. Neurol.* **7,** 294–298 (1994).
17. Ajami, B., Bennett, J. L., Krieger, C., Tetzlaff, W. & Rossi, F. M. Local self-renewal can sustain CNS microglia maintenance and function throughout adult life. *Nature Neurosci.* **10,** 1538–1543 (2007).
18. Bliss, T. V. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361,** 31–39 (1993).
19. Conboy, I. M. *et al.* Rejuvenation of aged progenitor cells by exposure to a young systemic environment. *Nature* **433,** 760–764 (2005).
20. Couillard-Despres, S. *et al.* In vivo optical imaging of neurogenesis: watching new neurons in the intact brain. *Mol. Imaging* **7,** 28–34 (2008).
21. Monje, M. L., Toda, H. & Palmer, T. D. Inflammatory blockade restores adult hippocampal neurogenesis. *Science* **302,** 1760–1765 (2003).
22. Moriyama, M. *et al.* Complement receptor 2 is expressed in neural progenitor cells and regulates adult hippocampal neurogenesis. *J. Neurosci.* **31,** 3981–3989 (2011).
23. Fumagalli, M. & d'Adda di Fagagna, F. SASPense and DDRama in cancer and ageing. *Nature Cell Biol.* **11,** 921–923 (2009).
24. Luo, J. *et al.* Glia-dependent TGF-β signaling, acting independently of the TH17 pathway, is critical for initiation of murine autoimmune encephalomyelitis. *J. Clin. Invest.* **117,** 3306–3315 (2007).
25. Xie, X. & Smart, T. G. Modulation of long-term potentiation in rat hippocampal pyramidal neurons by zinc. *Pflugers Arch.* **427,** 481–486 (1994).
26. Alamed, J., Wilcock, D. M., Diamond, D. M., Gordon, M. N. & Morgan, D. Two-day radial-arm water maze learning and memory task; robust resolution of amyloid-related memory deficits in transgenic mice. *Nature Protocols* **1,** 1671–1679 (2006).
27. Raber, J. *et al.* Irradiation enhances hippocampus-dependent cognition in mice deficient in extracellular superoxide dismutase. *Hippocampus* **21,** 72–80 (2011).
28. Brack, A. S. *et al.* Increased Wnt signaling during aging alters muscle stem cell fate and increases fibrosis. *Science* **317,** 807–810 (2007).

Author Contributions S.A.V. and T.W.-C. developed the concept and designed all experiments. S.A.V. and J.L. designed and performed *in vivo* experiments. S.A.V. performed behavioural experiments. K.I.M. assisted with surgery. B.Z. and X.S.X. performed electrophysiology. M.B. and A.E. analysed human data. G.B. assisted with fear conditioning and irradiation analysis. S.A.V., T.M.S. and J.-S.P. performed *in vitro* experiments. T.M.S. assisted with MCSF analysis. N.F. assisted with radial arm maze. Z.D. performed flow cytometry. K.M.L. performed irradiation. E.C. assisted with *in vivo* plasma experiments. D.R.G., G.L., E.R.P., J.A.K. and J.F.Q. identified aging subjects and provided human samples. S.C.-D. and L.A. provided reagents and mice. T.A.R. provided reagents, conceptual advice and edited the manuscript. S.A.V. collected data, performed data analysis and generated figures. S.A.V. and T.W.-C. wrote the manuscript. T.W.-C. supervised all aspects of this project. All authors had the opportunity to discuss results and comment on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.W.-C. (twc@stanford.edu).

METHODS

Mice. The following mouse lines were used: C57BL/6 (The Jackson Laboratory), C57BL/6 aged mice (National Institutes of Ageing), Dcx-Luc mice[20] and C57BL/6J-Act-GFP (Jackson Laboratory). For parabiosis experiments male and female C57BL/6 mouse cohorts were used. For all other *in vivo* pharmacological and behavioural studies young (2–3 months) wild-type C57BL/6 male mice were used. Mice were housed under specific pathogen-free conditions under a 12 h light–dark cycle and all animal handling and use was in accordance with institutional guidelines approved by the Veterans Affairs Palo Alto Committee on Animal Research.

Immunohistochemistry. Tissue processing and immunohistochemistry was performed on free-floating sections following standard published techniques[24]. Briefly, mice were anaesthetized with 400 mg kg^{-1} chloral hydrate (Sigma-Aldrich) and transcardially perfused with 0.9% saline. Brains were removed and fixed in phosphate-buffered 4% paraformaldehyde, pH 7.4, at 4 °C for 48 h before they were sunk through 30% sucrose for cryoprotection. Brains were then sectioned coronally at 40 μm with a cryomicrotome (Leica Camera) and stored in cryoprotective medium. Primary antibodies were: goat anti-Dcx (1:500; Santa Cruz Biotechnology), rat anti-BrdU (1:5,000; Accurate Chemical and Scientific Corp.), goat anti-Sox2 (1:200; Santa Cruz), mouse anti-NeuN (1:1,000; Chemicon), mouse anti-GFAP (1:1,500; DAKO) and mouse anti-CD68 (1:50; Serotec). After overnight incubation, primary antibody staining was revealed by biotinylated secondary antibodies and the ABC kit (Vector) with diaminobenzidine (DAB; Sigma-Aldrich) or fluorescence-conjugated secondary antibodies. For BrdU labelling, brain sections were pre-treated with 2 N HCl at 37 °C for 30 min before incubation with primary antibody. For double-label immunofluorescence of BrdU/NeuN or BrdU/GFAP, sections were incubated overnight with rat anti-BrdU, rinsed and incubated for 1 h with donkey anti-rat antibody (2.5 μg ml^{-1}; Vector) before they were stained with mouse anti-NeuN antibody. To estimate the total number of Dcx- or Sox2-positive cells per dentate gyrus immunopositive cells in the granule cell and subgranular cell layer of the dentate gyrus were counted in every sixth coronal hemibrain section through the hippocampus and multiplied by 12.

BrdU administration and quantification of BrdU-positive cells. 50 mg kg^{-1} of BrdU was injected intraperitoneally into mice once a day for 6 days, and mice were killed 28 days later or injected daily for 3 days before being killed. To estimate the total number of BrdU-positive cells in the brain, we performed DAB staining for BrdU on every sixth hemibrain section. The number of BrdU-positive cells in the granule cell and subgranular cell layer of the dentate gyrus were counted and multiplied by 12 to estimate the total number of BrdU-positive cells in the entire dentate gyrus. To determine the fate of dividing cells a total of 200 BrdU-positive cells across 4–6 sections per mouse were analysed by confocal microscopy for co-expression with NeuN and GFAP. The number of double-positive cells was expressed as a percentage of BrdU-positive cells.

Parabiosis and flow cytometry. Parabiosis surgery followed previously described procedures[19]. Pairs of mice were anaesthetized and prepared for surgery. Mirror-image incisions at the left and right flanks, respectively, were made through the skin. Shorter incisions were made through the abdominal wall. The peritoneal openings of the adjacent parabionts were sutured together. Elbow and knee joints from each parabiont were sutured together and the skin of each mouse was stapled (9-mm autoclip, Clay Adams) to the skin of the adjacent parabiont. Each mouse was injected subcutaneously with Baytril antibiotic and Buprenex as directed for pain and monitored during recovery. Flow cytometric analysis was done on fixed and permeabilized blood plasma cells from GFP and non-GFP parabionts. Approximately 40–60% of cells in the blood of either parabiont were GFP-positive 2 weeks after parabiosis surgery. We observed 70–80% survival rate in parabionts 5 weeks after parabiosis surgery.

Extracellular electrophysiology. Acute hippocampal slices (400-μm thick) were prepared from unpaired and young parabionts. Slices were maintained in artificial cerebrospinal fluid (ACSF) continuously oxygenated with 5% CO$_2$/95% O$_2$. ACSF composition was as follows: (in mM): NaCl 124.0; KCl 2.5; KH$_2$PO$_4$ 1.2; CaCl$_2$ 2.4; MgSO$_4$ 1.3; NaHCO$_3$ 26.0; glucose 10.0 (pH 7.4). Recordings were performed with an Axopatch- 2B amplifier and pClamp 10.2 software (Axon Instruments). Submerged slices were continuously perfused with oxygenated ACSF at a flow rate of 2 ml min^{-1} from a reservoir by gravity feeding. Field potential (population spikes and EPSPs) was recorded using glass microelectrodes filled with ACSF (resistance: 4–8 MΩ). Biphasic current pulses (0.2 ms duration for one phase, 0.4 ms in total) were delivered in 10-s intervals through a concentric bipolar stimulating electrode (FHC). No obvious synaptic depression or facilitation was observed with this frequency stimulation. To record field population spikes in the dentate gyrus, the recording electrode was placed in the lateral or medial side of the dorsal part of the dentate gyrus. The stimulating electrode was placed right above the hippocampal fissure to stimulate the perforant pathway fibres. Signals were filtered at 1 KHz and digitized at 10 KHz. Tetanic stimulation consisted of 2 trains of 100 pulses (0.4 ms pulse duration, 100 Hz) delivered with an inter-train interval

of 5 s. The amplitude of the population spike was measured from the initial phase of the negative wave. Up to five consecutive traces were averaged for each measurement. Synaptic transmission was assessed by generating input–output curves, with stimulus strength adjusted to be ~30% of the maximum. LTP was calculated as mean percentage change in the amplitude of the population spike following high-frequency stimulation relative to its basal amplitude.

Contextual fear conditioning. The paradigm was done following previously published techniques[27]. In this task, mice learned to associate the environmental context (fear-conditioning chamber) with an aversive stimulus (mild foot shock; unconditioned stimulus (US)), enabling testing for hippocampal-dependent contextual fear conditioning. As contextual fear conditioning is hippocampus and amygdala dependent, the mild foot shock was paired with a light and tone cue (conditioned stimulus (CS)) in order to also assess amygdala-dependent cued fear conditioning. Conditioned fear was displayed as freezing behaviour. Specific training parameters are as follows: tone duration is 30 s; level is 70 dB, 2 kHz; shock duration is 2 s; intensity is 0.6 mA. This intensity is not painful and can easily be tolerated but will generate an unpleasant feeling. More specifically, on day 1 each mouse was placed in a fear-conditioning chamber and allowed to explore for 2 min before delivery of a 30 s tone (70 dB) ending with a 2 s foot shock (0.6 mA). Two minutes later, a second CS–US pair was delivered. On day 2 each mouse was first place in the fear-conditioning chamber containing the same exact context, but with no adminstration of a CS or foot shock. Freezing was analysed for 1–3 min. One hour later, the mice were placed in a new context containing a different odour, cleaning solution, floor texture, chamber walls and shape. Animals were allowed to explore for 2 min before being re-exposed to the CS. Freezing was analysed for 1–3 min. Freezing was measured using a FreezeScan video tracking system and software (Cleversys).

RAWM. Spatial learning and memory was assessed using the RAWM paradigm following the exact protocol described previously[26]. The goal arm location containing a platform remains constant throughout the training and testing phase, whereas the start arm is changed during each trial. On day 1 during the training phase, mice are trained for 15 trials, with trials alternating between a visible and hidden platform. On day 2 during the testing phase, mice are tested for 15 trials with a hidden platform. Entry into an incorrect arm is scored as an error, and errors are averaged over training blocks (three consecutive trials).

Cranial irradiation. Adult mice (8–12 weeks) were sham irradiated (controls) or irradiated at 5 Gy three times over 8 days using the Mark I gamma irradiator and killed at 8–10 weeks after irradiation to collect brains for immunohistochemical analyses. Each mouse was placed in a restrainer that was fitted into a slot in the lead brick shield so that the back of the skull was facing the source of radiation when positioned in the radiation chamber. The shield is constructed of lead bricks such that only the hippocampal/midbrain area was exposed to radiation. Calibration for 5 Gy radiation was done using nanoDot. Shielded areas were protected with an exposure rate ten times lower than the exposed area. RAWM studies were done on irradiated mice at least 6 weeks after the radiation procedure. This time frame ensured adequate recovery of the animals. All data were from 8 irradiated and 10 sham-irradiated mice.

Plasma collection and proteomic analysis. Mouse blood was collected from 400–500 young (2–3 months) and old (18–22 months) animals into EDTA-coated tubes via tail vein bleed, mandibular vein bleed, or intracardial bleed at time of death. EDTA plasma was generated by centrifugation of freshly collected blood and aliquots were stored at −80 °C until use. Human plasma and CSF samples were obtained from academic centres and subjects were chosen based on standardized inclusion and exclusion criteria as previously described[29,30]. The relative plasma concentrations of cytokines and signalling molecules were measured in human and mouse plasma samples using standard antibody-based multiplex immunoassays (Luminex) by either Rules Based Medicine, a fee-for-service provider, or by the Human Immune Monitoring Center at Stanford University. All Luminex measurements where obtained in a blinded fashion. All assays were developed and validated to Clinical Laboratory Standards Institute (formerly NCCLS) guidelines based upon the principles of immunoassay as described by the manufacturers.

CCL11, MSCF, antibody, or plasma administration. Carrier-free recombinant murine CCL11 dissolved in PBS (10 μg kg^{-1}; R&D Systems), carrier-free recombinant MCSF dissolved in PBS (10 μg kg^{-1}; Biogen), rat IgG2a neutralizing antibody against mouse CCL11 (50 μg kg^{-1}; R&D Systems, clone 42285), and isotype-matched control rat IgG2a recommended by the manufacturer (R&D Systems, clone 54447) were administered systemically via intraperitoneal injection over ten days on day 1, 4, 7 and 10. The same reagents (0.50 μl; 0.1 μg μl^{-1}) were also administered stereotaxically into the dentate gyrus of the hippocampus in some experiments (coordinates from bregma: $A = -2.0$ mm and $L = -1.8$ mm; from brain surface: $H = -2.0$ mm). Pooled mouse serum or plasma was collected from young (2–3 months) mice and old (18–20 months) mice by intracardial bleed at time of death. Serum was prepared from clotted blood collected without anticoagulants;

plasma was prepared from blood collected with EDTA followed by centrifugation. Aliquots were stored at $-80\,^{\circ}$C until use. Prior to administration plasma was dialysed in PBS to remove EDTA. Young adult mice were systemically treated with plasma ($100\,\mu$l) isolated from young or aged mice via intravenous injections four times over 10 days.

***In vivo* bioluminescence imaging.** Bioluminescence was detected with the *In vivo* Imaging System (IVIS Spectrum; Caliper Life Science). Mice were injected intraperitoneally with $150\,mg\,kg^{-1}$ D-luciferin (Xenogen) 10 min before imaging and anaesthetized with isofluorane during imaging. Photons emitted from living mice were acquired as photons per second per cm^2 per steridan (sr) using LIVINGIMAGE software (version 3.5, Caliper) and integrated over 5 min. For quantification a region of interest was manually selected and kept constant for all experiments.

Cell culture assays. Mouse NPCs were isolated from C57BL/6 mice as previously described[12]. Brains from postnatal animals (1-day old) were dissected to remove the olfactory bulb, cerebellum and brainstem. After removing superficial blood vessels forebrains were finely minced, digested for 30 min at $37\,^{\circ}$C in DMEM media containing $2.5\,U\,ml^{-1}$ Papain (Worthington Biochemicals), $1\,U\,ml^{-1}$ Dispase II (Boeringher Mannheim) and $250\,U\,ml^{-1}$ DNase I (Worthington Biochemicals) and mechanically dissociated. NPCs were purified using a 65% Percoll gradient and plated on uncoated tissue culture dishes at a density of $10^5\,cells\,cm^{-2}$. NPCs were cultured under standard conditions in NeuroBasal A medium supplemented with penicillin ($100\,U\,ml^{-1}$), streptomycin ($100\,mg\,ml^{-1}$), $2\,mM$ L-glutamine, serum-free B27 supplement without vitamin A (Sigma-Aldrich), bFGF ($20\,ng\,ml^{-1}$) and EGF ($20\,ng\,ml^{-1}$). Carrier-free forms of murine recombinant CCL11 ($100\,ng\,ml^{-1}$, R&D Systems), goat IgG neutralizing antibody against mouse CCL11 ($10\,\mu g\,ml^{-1}$; R&D Systems), and control goat IgG ($10\,\mu g$ ml^{-1}; R&D Systems) were dissolved in PBS and added to cell cultures under self-renewal conditions every other day following cell plating.

Human NTERA cells[7] expressing eGFP under the Dcx promoter were cultured under standard self-renewal and differentiation conditions[31,32]. Carrier-free forms of human recombinant CCL11 ($100\,ng\,ml^{-1}$, R&D Systems), mouse IgG$_1$ neutralizing antibody against human CCL11 ($25\,\mu g\,ml^{-1}$; R&D Systems, clone 43911) and control mouse IgG$_1$ ($25\,\mu g\,ml^{-1}$; R&D Systems) were added to cell cultures under differentiation conditions every other day following cell plating.

Data and statistical analysis. Data are expressed as mean $\pm$ s.e.m. Statistical analysis was performed with Prism 5.0 software (GraphPad Software). Means between two groups were compared with two-tailed, unpaired Student's *t*-test. Comparisons of means from multiple groups with each other or against one control group were analysed with one-way ANOVA and Tukey-Kramer's or Dunnett's post-hoc tests, respectively. Plasma protein correlations in the ageing samples were analysed with the Significance Analysis of Microarray software (*SAM 3.00* algorithm; http://www.stat.stanford.edu/~tibs/SAM/index.htm). Unsupervised cluster analysis was performed using Gene Cluster 3.0 software and node maps were produced using Java TreeView 1.0.13 software. All histology, electrophysiology and behaviour experiments conducted were done in a randomized and blinded fashion.

29. Zhang, J. *et al.* CSF multianalyte profile distinguishes Alzheimer and Parkinson diseases. *Am. J. Clin. Pathol.* **129**, 526–529 (2008).
30. Li, G. *et al.* Cerebrospinal fluid concentration of brain-derived neurotrophic factor and cognitive function in non-demented subjects. *PLoS ONE* **4**, e5424 (2009).
31. Couillard-Despres, S. *et al.* Human *in vitro* reporter model of neuronal development and early differentiation processes. *BMC Neurosci.* **9**, 31 (2008).
32. Buckwalter, M. S. *et al.* Chronically increased transforming growth factor-β1 strongly inhibits hippocampal neurogenesis in aged mice. *Am. J. Pathol.* **169**, 154–164 (2006).

LETTER

Cell–to–cell spread of HIV permits ongoing replication despite antiretroviral therapy

Alex Sigal[1], Jocelyn T. Kim[1,2], Alejandro B. Balazs[1], Erez Dekel[3], Avi Mayo[3], Ron Milo[4] & David Baltimore[1]

Latency and ongoing replication[1] have both been proposed to explain the drug-insensitive human immunodeficiency virus (HIV) reservoir maintained during antiretroviral therapy. Here we explore a novel mechanism for ongoing HIV replication in the face of antiretroviral drugs. We propose a model whereby multiple infections[2,3] per cell lead to reduced sensitivity to drugs without requiring drug-resistant mutations, and experimentally validate the model using multiple infections per cell by cell-free HIV in the presence of the drug tenofovir. We then examine the drug sensitivity of cell-to-cell spread of HIV[4–7], a mode of HIV transmission that can lead to multiple infection events per target cell[8–10]. Infections originating from cell-free virus decrease strongly in the presence of antiretrovirals tenofovir and efavirenz whereas infections involving cell-to-cell spread are markedly less sensitive to the drugs. The reduction in sensitivity is sufficient to keep multiple rounds of infection from terminating in the presence of drugs. We examine replication from cell-to-cell spread in the presence of clinical drug concentrations using a stochastic infection model and find that replication is intermittent, without substantial accumulation of mutations. If cell-to-cell spread has the same properties *in vivo*, it may have adverse consequences for the immune system[11–13], lead to therapy failure in individuals with risk factors[14], and potentially contribute to viral persistence and hence be a barrier to curing HIV infection.

Current antiretroviral therapy (ART) does not cure HIV infection because low-level viraemia persists from virus reservoirs that are insensitive to ART[1]. The reservoirs may be long-lived infected cells, cells with latent virus, ongoing cycles of infection termed ongoing replication, or a combination of sources[1]. How ongoing replication might take place in the face of ART has remained unclear. If ART succeeds in decreasing ongoing HIV replication to very low levels, why does it not eliminate replication completely? Here we explore a novel mechanism for ongoing HIV replication in the presence of ART.

Multiple infections of one cell may propagate at drug concentrations where infection by single particles would die out: if more virions are transmitted per cell, the probability that at least one of the virions escapes the drug should increase (Fig. 1a). To model the effect of multiple infections on drug sensitivity (Supplementary Theory, section 1), we assume infections by individual virions are independent events, each with a probability of escaping the drug and succeeding in infecting the cell. To quantify infection sensitivity to drugs, we introduce the transmission index (T_X), which we define as the fraction of cells infected in the presence of drug (I_d) divided by the fraction of cells infected in the absence of drug (I). Given: (1) a multiplicity of infection of m infectious units per cell, where m is defined as the product of virus particle number and the probability of infection per virus particle; (2) a concentration of antiretroviral agent d that reduces m by factor $f(d)$, where $f(d) \geq 1$. Under these conditions, the transmission index is:

$$T_X = \frac{I_d}{I} = \frac{1 - e^{-m/f(d)}}{1 - e^{-m}} \quad (1)$$

T_X has two important limiting regimes: $m \ll 1$, in which case $T_X \approx 1/f(d)$ and $m/f(d) \gg 1$, in which case $T_X \approx 1$. In the first case, where few viruses infect each cell, the infection is sensitive to the effect of the drug, whereas in the second, where many viruses infect each cell, the infection is insensitive.

To test this, we infected the highly infection-permissive MT-4 T-cell line with cell-free HIV encoding yellow fluorescence protein (YFP)[15] at low (0.2) and high (100) m in the presence of tenofovir (TFV), a nucleotide reverse transcriptase inhibitor. We determined infected cell number by YFP fluorescence (Supplementary Fig. 1) and observed that infection with cell-free virus at low m was sensitive to TFV across the range of concentrations used. At high m, infection was insensitive to low and intermediate TFV concentrations (Fig. 1b), supporting the model. Thus, multiple cell-free HIV infections per cell recapitulate the insensitivity to drug of an HIV reservoir.

Multiple infections occur *in vivo*[2,16] and in culture[8,10] and are thought to be associated with cell-to-cell spread[2,8–10], a directed transmission mode that minimizes the number of virus particles failing to
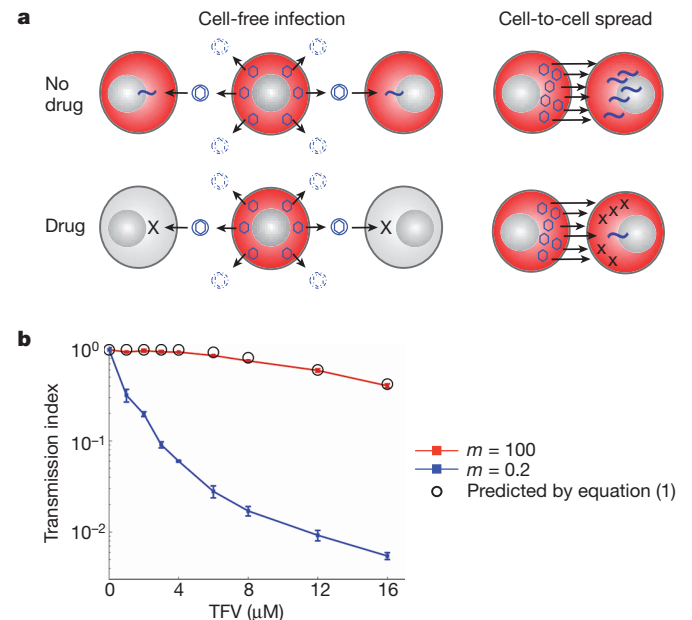
Figure 1 | Multiple infections per cell decrease sensitivity to drug.
a, Hypothesis. Red circles indicate infected cells, arrows indicate transmissions, hexagons or hexagons surrounded by circles indicate viruses, broken circles indicate degraded viruses, crosses indicate viruses blocked by drug and wavelets indicate successful infection. **b**, MT-4 cells were pre-incubated with TFV and infected with HIV coding for YFP. Infection multiplicity m was 0.2 (blue squares) or 100 (red squares). Lines are a guide for the eye. Mean ± standard deviation (s.d.) of replicates ($n = 3$). Circles represent calculated values of T_X at $m = 100$ according to equation (1) with $f(d)$ at each drug concentration determined empirically at $m = 0.2$.

[1]Division of Biology, California Institute of Technology, Pasadena, California 91125, USA. [2]Division of Infectious Diseases, Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California 90095, USA. [3]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel. [4]Department of Plant Sciences, Weizmann Institute of Science, Rehovot 76100, Israel.

reach the target cell. We therefore used co-culture with infected cells to generate cell-to-cell spread and compared drug sensitivity to infection with cell-free virus. Infection by co-culture occurs both by cell-free virus shed by infected donor cells and by cell-to-cell spread. Administration of cell-free virus lacks a cell-to-cell component—the measured average virus cycle time (1.7 days; Supplementary Fig. 2) would rarely permit cell-free virus infected cells to complete a second round of infection during the experiment (2 days). Therefore, we compared cell-free virus infection and the combination of cell-free virus infection and cell-to-cell spread resulting from co-culture. We used drugs that act far downstream of entry, to ensure any differences between cell-to-cell and cell-free infection are not due to factors that physically inhibit drug action in cell-to-cell spread.

We infected peripheral blood mononuclear cells (PBMCs) in the presence or absence of TFV by co-culture or using cell-free virus. To separate donor from target cells in co-culture, we used HLA-A2-negative donor cells and HLA-A2-positive targets (Supplementary Fig. 3a). Two days post-infection, we determined the fraction of target cells infected using p24 intracellular staining of HLA-A2-positive PBMCs (Fig. 2a, top panel, controls in Supplementary Fig. 3b). Co-culture dramatically decreased sensitivity to drug: TFV decreased cell-free infection ~30-fold but caused less than a twofold decrease of co-culture infection (Fig. 2b). The decline in HLA-A2 expression in the target cells after infection (Supplementary Fig. 3b) is consistent with observations that productive HIV infection downregulates HLA[17].

We also used Rev-CEM[18] reporter T cells as targets. These cells express green fluorescent protein (GFP) in the presence of HIV early proteins Tat and Rev (Supplementary Fig. 4). To infect Rev-CEM cells, we used either cell-free HIV or co-culture with infected MT-4 cells engineered to be >99% mCherry positive (Supplementary Fig. 5). We excluded GFP/mCherry double-positive cells from the analysis to avoid scoring fused cells as infected (Supplementary Figs 6 and 7). This underestimates co-culture infection because it excludes unfused cell doublets in the process of virus exchange.

To block infection, we applied TFV and the non-nucleoside reverse transcriptase inhibitor efavirenz (EFV) (Fig. 2a, bottom panel, Supplementary Fig. 7). At the highest concentrations used, co-culture T_X was over sixfold higher than cell-free infection T_X (Fig. 2b). The trend was similar when donors were PBMCs or Rev-CEM cells (Supplementary Fig. 8). Co-culture T_X was lower than in PBMC-to-PBMC transmission, suggesting that target cells have an important role in cell-to-cell spread efficiency. The lower drug sensitivity in co-culture was not due to secreted donor cell factors that decrease the susceptibility of target cells to drugs (Supplementary Fig. 9).

We next determined the number of infectious units (m) transmitted. For co-culture, m was previously proposed to have a two-peaked Poisson distribution, one peak corresponding to cell-free virus or some low virus cell-to-cell transmissions, and the second to high virus number transmissions[3,9]. We fit a two-peaked Poisson and other distributions to the data (Supplementary Theory, section 2). The two-peaked Poisson fit the data best (Fig. 2b, dotted line, Supplementary Fig. 10). The first peak mean was ~1 infectious unit for both drugs, with 94% and 97% of infections in this peak for TFV and EFV, respectively. The second peak mean was 73 (TFV) and 175 (EFV), with the remaining 6% and 3% of infections in this peak. This predicts that whereas most infections are cell-free or low virus cell-to-cell transmissions, a minority involve very large numbers of viruses. This might seem to imply large numbers of integrations in the absence of drug in the high virus number subset. Arguing against this is our observation of a significantly increased cell death rate with increasing numbers of multiple infections in the absence of drugs (data not shown). Inter-virus interference, such as downregulation of CD4 receptors[19], may also limit provirus number.
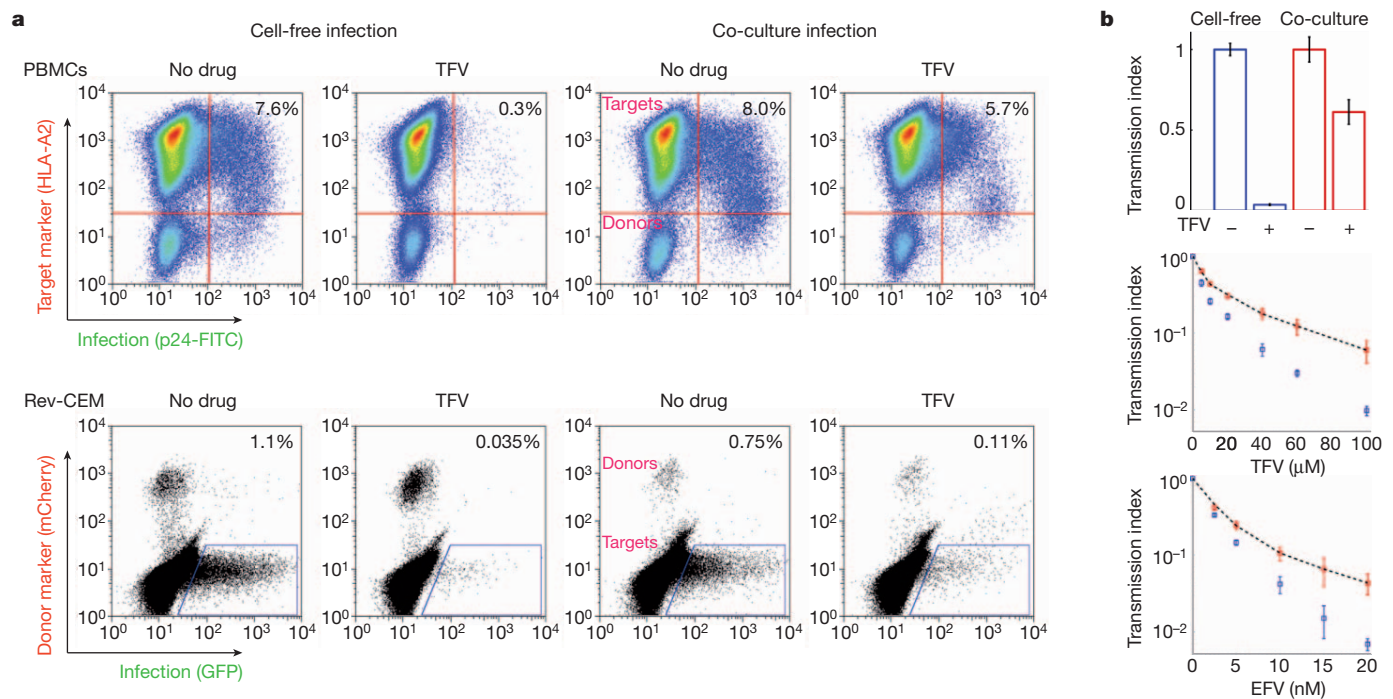
Figure 2 | Cell-to-cell spread reduces sensitivity to drugs. a, Top, infection of HLA-A2-positive PBMC targets with cell-free virus (left two plots) or infected HLA-A2-negative PBMC donors (right two plots) in the absence or presence of 10 μM TFV. x-axis is p24, y-axis HLA-A2 status. Bottom, the number of GFP-positive Rev-CEM cells after infection with cell-free virus (left two plots) or infected MT-4mCherry donors (right two plots) in the absence or presence of 60 μM TFV. x-axis is GFP, y-axis is mCherry fluorescence. **b**, Transmission index when infection source was cell-free HIV (blue bars or squares) or co-culture with HIV-infected donor cells (red bars or squares). Mean ± s.d.

($n = 3$). Top graph is PBMCs with TFV, middle graph is Rev-CEM cells with TFV, bottom graph is Rev-CEM cells with EFV. Black dashed line is best fit of m with a two-peaked Poisson distribution described by
$$p(m; a, \mu_1, \mu_2) = (1-a)e^{-\mu_1}\mu_1^m/m! + ae^{-\mu_2}\mu_2^m/m!,$$ where μ_1 and μ_2 are the means of the first and second peak respectively, and a is the fraction of transmissions that fall within the second peak. For TFV, $\mu_1 = 1.1$, $\mu_2 = 73$, $a = 0.06$. For EFV, $\mu_1 = 0.8$, $\mu_2 = 175$, with $a = 0.03$. Root mean squared error was 0.01 (EFV) to 0.02 (TFV).

To investigate whether cell-to-cell spread can lead to HIV replication through multiple virus cycles with ART, we measured the replication ratio (R), defined as fold change in the number of infected cells per virus cycle under conditions where target cells are not limiting: $(I_k/I_0)^{1/k}$. Here k is the number of elapsed virus cycles, I_k is the number of infected cells at virus cycle k, and I_0 is the number of infected cells at the start. For expanding infections $R > 1$, whereas infections with $R < 1$ ultimately terminate[20,21]. Although this assumes synchronized virus cycles, we simulated desynchronization and observed that its effect was negligible at the measured variability in cycle lengths (Supplementary Fig. 11).

To measure R, we tracked infection daily (Methods) in the absence of drug, with 100 μM TFV, or with a combination of EFV, TFV and the nucleoside reverse transcriptase inhibitor emtricitabine (FTC) at their clinical maximum plasma concentrations (C_{max}: 10 μM EFV, 2 μM TFV and 10 μM FTC[22]). The fraction of infected cells was kept low to ensure that target cells were not limiting. R_0, R_{TFV} and $R_{C_{max}}$, the replication ratios with no drug, TFV or at C_{max}, were fitted from the data (Fig. 3a, dashed lines). They were 65, 2.5 and 0.95, respectively. R_{TFV} was significantly greater than 1 ($P < 0.01$), indicating an expanding infection. $R_{C_{max}}$ was slightly lower than 1 in all experiments (Supplementary Fig. 12), indicating an infection slightly below the expansion threshold.

We compared experimentally obtained replication ratios with those predicted for the same drug concentrations if cell-free infection were the only infection route (Supplementary Theory section 3 and Supplementary Fig. 13). We obtained $R_{TFV} = 1.1$ and $R_{C_{max}} = 0.60$ values in this case (Fig. 3a). The predicted R with no replication, resulting solely from infected cell half-life, was 0.46 (Fig. 3a). Predicted cell-free replication ratios were significantly lower ($P < 0.02$ for TFV, $P < 0.01$ for C_{max}) than ratios experimentally obtained from co-culture.

Given the lack of evolution in the plasma in individuals with HIV successfully suppressed by drugs[23,24], ongoing replication can occur if:

(1) it is compartmentalized to other locations[25,26], (2) if it is intermittent; (3) the circulating virus is at a fitness maximum[24]; or some combination of these factors. We obtained $R_{C_{max}} = 0.95$. If this is extrapolated *in vivo*, it follows that ongoing replication cannot persist independently but may have a role if it interacts with another reservoir that primes replication[27]. To examine this scenario, we performed a stochastic simulation (Methods). As expected for intermittent replication, every infection chain that starts from the introduction of an infected cell from a different reservoir—for example, reactivation from latency—terminates (Supplementary Fig. 14). A constant input of one infected cell per virus cycle results in a steady state where substantial numbers of newly infected cells are generated, but the average number of mutations anywhere on the HIV genome per infected cell is low (~1; Fig. 3b). Because each infection chain is independent, these mutations are expected to be sporadic and not linked by temporal structure.

Evidence for ongoing replication during ART derives from the decrease in virus decline rates[28], some HIV sequence divergence[29] and long terminal repeat circle formation when the integrase inhibitor raltegravir is included in drug regimens[11]. At least in some individuals, antiretroviral suppression is close to the ongoing replication threshold: a mutation conferring very low-level resistance to EFV at therapy initiation[30] is sufficient to cause ongoing replication, as indicated by increased virological failure risk[14]. Our data indicate that cell-to-cell spread is a likely source of intermittent ongoing replication in the face of ART, and that this is a consequence of some cell-to-cell infections transmitting virus numbers much in excess of what is required to infect a cell in the absence of ART. The large transmitted dose strongly decreases the probability that every transmitted virus will be inhibited by the drugs, and therefore greatly weakens their effect. This replication may adversely affect the immune system, increasing activation[11,12] and cell death[13], and could potentially contribute to the maintenance of an HIV reservoir in locations such as lymphoid tissue where cell-to-cell spread occurs.

METHODS SUMMARY

HIV infection at high and low m. NL4-3YFP HIV stock at a 1:2,000 or 1:4 final dilution was added to MT-4 cells pre-incubated with TFV. Two days post-infection, the number of YFP-positive cells was determined by FACS. The multiplicity of infection was calculated as $m = -\ln p(0) = -\ln(1 - I_{1:2,000})$, where $I_{1:2,000}$ is the fraction of YFP-positive cells at the 1:2,000 dilution.

Drug sensitivity of co-culture versus cell-free infections. Donor cells were infected with NL4-3 strain HIV and incubated for two to three days. Infected donor cells or cell-free NL4-3 were then added to target cells. Two days after target cell infection, the number of infected cells was determined by FACS using intracellular p24 staining (PBMCs) or GFP expression (Rev-CEM cells). In all experiments, uninfected PBMC or MT-4mCherry cells were added to cell-free virus infections to keep total cell numbers equal on day 0.

Infection growth rate. Infection was initiated by adding infected Rev-CEM cells to uninfected Rev-CEM cells pre-incubated with drugs. Cells were passaged on each day following infected cell addition: infection with no drug was split 1:10 into fresh Rev-CEM cells. For 100 μM TFV or C_{max}, infected cells were split 0.6:1 with drug-containing medium. Cells remaining after split were used to quantify the fraction of infected cells by FACS. The fold change in infected cells on each day was calculated as $N_k D_k/N_0$, where N_k is the fraction of infected cells on day k, D_k is total dilution factor (split) up to day k and N_0 is the fraction of infected cells on day 1.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Figure 3 | Co-culture infection dynamics. a, Infection growth rate. Drug conditions were: no drug (black squares), 100 μM TFV (red squares) and C_{max} (blue squares). Means ± s.d. of inter-day experiments ($n = 3$). Dashed lines represent fits of $I_k = I_0 R^k$ for each drug condition. Solid lines are predicted infection dynamics for infection occurring exclusively by cell-free virus in the presence of 100 μM TFV (red line), C_{max} (blue line), or with no viral replication (green line). **b,** Simulation of the number of infected cells and mutations per cell with an input of one infected cell per virus cycle (Methods). x-axis is time, y-axis is number of total infected cells (top graph), newly infected cells (middle graph) or sum of mutations divided by the sum of total infected cells (bottom graph). Average number of mutations per cell over time is 0.6 ± 0.5 (mean ± s.d., $n = 215$).

1. Pierson, T., McArthur, J. & Siliciano, R. F. Reservoirs for HIV-1: mechanisms for viral persistence in the presence of antiviral immune responses and antiretroviral therapy. *Annu. Rev. Immunol.* **18**, 665–708 (2000).
2. Jung, A. *et al.* Recombination: multiply infected spleen cells in HIV patients. *Nature* **418**, 144 (2002).
3. Dixit, N. M. & Perelson, A. S. HIV dynamics with multiple infections of target cells. *Proc. Natl Acad. Sci. USA* **102**, 8198–8203 (2005).
4. Dimitrov, D. S. *et al.* Quantitation of human immunodeficiency virus type 1 infection kinetics. *J. Virol.* **67**, 2182–2190 (1993).

5. Sattentau, Q. Avoiding the void: cell-to-cell spread of human viruses. *Nature Rev. Microbiol.* **6,** 815–826 (2008).
6. Martin, N. *et al.* Virological synapse-mediated spread of human immunodeficiency virus type 1 between T cells is sensitive to entry inhibition. *J. Virol.* **84,** 3516–3527 (2010).
7. Sourisseau, M., Sol-Foulon, N., Porrot, F., Blanchet, F. & Schwartz, O. Inefficient human immunodeficiency virus replication in mobile lymphocytes. *J. Virol.* **81,** 1000–1012 (2007).
8. Dang, Q. *et al.* Nonrandom HIV-1 infection and double infection via direct and cell-mediated pathways. *Proc. Natl Acad. Sci. USA* **101,** 632–637 (2004).
9. Dixit, N. M. & Perelson, A. S. Multiplicity of human immunodeficiency virus infections in lymphoid tissue. *J. Virol.* **78,** 8942–8945 (2004).
10. Del Portillo, A. *et al.* Multiploid inheritance of HIV-1 during cell-to-cell infection. *J. Virol.* **85,** 7169–7176 (2011).
11. Buzón, M. J. *et al.* HIV-1 replication and immune dynamics are affected by raltegravir intensification of HAART-suppressed subjects. *Nature Med.* **16,** 460–465 (2010).
12. Chun, T. W. *et al.* Relationship between residual plasma viremia and the size of HIV proviral DNA reservoirs in infected individuals receiving effective antiretroviral therapy. *J. Infect. Dis.* **204,** 135–138 (2011).
13. Doitsh, G. *et al.* Abortive HIV infection mediates CD4 T cell depletion and inflammation in human lymphoid tissue. *Cell* **143,** 789–801 (2010).
14. Paredes, R. *et al.* Pre-existing minority drug-resistant HIV-1 variants, adherence, and risk of antiretroviral treatment failure. *J. Infect. Dis.* **201,** 662–671 (2010).
15. Levy, D. N., Aldrovandi, G. M., Kutsch, O. & Shaw, G. M. Dynamics of HIV-1 recombination in its natural target cells. *Proc. Natl Acad. Sci. USA* **101,** 4204–4209 (2004).
16. Gratton, S., Cheynier, R., Dumaurier, M. J., Oksenhendler, E. & Wain-Hobson, S. Highly restricted spread of HIV-1 and multiply infected cells within splenic germinal centers. *Proc. Natl Acad. Sci. USA* **97,** 14566–14571 (2000).
17. Collins, K. L., Chen, B. K., Kalams, S. A., Walker, B. D. & Baltimore, D. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature* **391,** 397–401 (1998).
18. Wu, Y., Beddall, M. H. & Marsh, J. W. Rev-dependent indicator T cell line. *Curr. HIV Res.* **5,** 394–402 (2007).
19. Chen, B. K., Gandhi, R. T. & Baltimore, D. CD4 down-modulation during infection of human T cells with human immunodeficiency virus type 1 involves independent activities of vpu, env, and nef. *J. Virol.* **70,** 6044–6053 (1996).
20. Ribeiro, R. M. *et al.* Estimation of the initial viral growth rate and basic reproductive number during acute HIV-1 infection. *J. Virol.* **84,** 6096–6102 (2010).
21. Nowak, M. A. & May, R. M. *Virus Dynamics: Mathematical Principles of Immunology and Virology* (Oxford Univ. Press, 2000).
22. Mathias, A. A. *et al.* Bioequivalence of efavirenz/emtricitabine/tenofovir disoproxil fumarate single-tablet regimen. *J. Acquir. Immune Defic. Syndr.* **46,** 167–173 (2007).
23. Bailey, J. R. *et al.* Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4+ T cells. *J. Virol.* **80,** 6441–6457 (2006).
24. Kieffer, T. L. *et al.* Genotypic analysis of HIV-1 drug resistance at the limit of detection: virus production without evolution in treated adults with undetectable HIV loads. *J. Infect. Dis.* **189,** 1452–1465 (2004).
25. Cu-Uvin, S. *et al.* Genital tract HIV-1 RNA shedding among women with below detectable plasma viral load. *AIDS* **24,** 2489–2497 (2010).
26. North, T. W. *et al.* Viral sanctuaries during highly active antiretroviral therapy in a nonhuman primate model for AIDS. *J. Virol.* **84,** 2913–2922 (2010).
27. Rong, L. & Perelson, A. S. Modeling latently infected cell activation: viral and latent reservoir persistence, and viral blips in HIV-infected patients on potent therapy. *PLOS Comput. Biol.* **5,** e1000533 (2009).
28. Grossman, Z. *et al.* Ongoing HIV dissemination during HAART. *Nature Med.* **5,** 1099–1104 (1999).
29. Frenkel, L. M. *et al.* Multiple viral genetic analyses detect low-level human immunodeficiency virus type 1 replication during effective highly active antiretroviral therapy. *J. Virol.* **77,** 5721–5730 (2003).
30. Rhee, S. Y. *et al.* Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* **31,** 298–303 (2003).

Author Contributions A.S. and D.B. conceived the study. A.S. designed the research; A.S. and J.T.K. performed the experiments with support from A.B.B.; A.S. formulated the basic mathematical model and performed the numerical simulations; R.M., A.M. and E.D. added analytical insights and expanded the model to treat virus number as a random variable; A.S. and D.B. wrote the paper.

METHODS

Cells, viruses and drugs. The following were obtained through the AIDS Research and Reference Reagent Program, National Institute of Allergy and Infectious Diseases, National Institutes of Health: Rev-CEM cells from Y. Wu and J. Marsh; MT-4 cells from D. Richman; HIV expression plasmid pNL4-3 from M. Martin; TFV; EFV. The NL4-3YFP molecular clone was a gift from D. Levy. Cell-free virus was produced by transfection of HEK293 cells with virus coding plasmid using Fugene 6 or Fugene HD (Roche). Supernatant containing shed virus was harvested after two days of incubation. Number of virus genomes of viral stock was determined using the RealTime HIV-1 Viral Load test (Abbott Diagnostics, Abbott Park Ill) and gag p24 content was determined by ELISA (Perkin-Elmer) at the the ARI-UCSF Laboratory of Clinical Virology. The MT-4mCherry cell line was created by infecting MT-4 cells with a pHAGE2 lentiviral vector expressing mCherry under the control of the EF1α promoter. To obtain a 0.99 fraction of mCherry-positive cells, MT-4mCherry cells were used fresh after lentiviral infection without a cycle of freezing and thawing, minimizing the number of population doublings and consequent decrease in the mCherry-positive fraction. Anonymous PBMCs or peripheral blood samples were provided by AllCells (PBMCs) or the UCLA Center for AIDS Research (CFAR) Virology Core Lab (peripheral blood). For whole blood, PBMCs were purified by Ficoll gradient using standard techniques. Purified PBMCs were activated with $5\,\mu g\,ml^{-1}$ PHA in the presence of $5\,ng\,ml^{-1}$ IL-2 for 1 (donors) or 3 (targets) days. All work was approved by the California Institute of Technology Institutional Biosafety Committee and Institutional Review Board exempt.

NL4-3YFP infection at high and low m. MT-4 cells were pre-incubated for 24 h with varying concentrations of TFV. NL4-3YFP stock was produced using transfection of HEK293 cells at 80% confluence with Fugene HD. Virus supernatant was collected 2 days post-transfection and added fresh to maximize the number of infectious units. Fresh virus stock was used at a 1:2,000 (low m) or 1:4 final dilution (high m). After 2 days incubation with virus, the number of YFP-positive MT-4 cells was quantified by flow cytometry by collecting 2×10^5 cells using a FACScaliber machine (Becton Dickenson). The multiplicity of infection was calculated using Poisson statistics: $m = -\ln p(0) = -\ln(1 - I_{1:2,000})$, where $p(0)$ is the fraction of YFP-negative cells, and $I_{1:2,000}$ is the fraction of YFP-positive cells at the 1:2,000 dilution.

Comparison of co-culture and cell-free infections in PBMCs. For PBMC infections, 1.5×10^6 PHA-activated HLA-A2-negative donor PBMCs at 10^6 cells ml^{-1} were either infected with 700 ng HIV (NL4-3 strain), or mock infected with the same volume of growth medium. Cells were then incubated for 2 days. Two days after donor-cell infection, PHA-activated HLA-A2-positive PBMC target cells at 10^6 cells ml^{-1} were either treated with no drug or $10\,\mu M$ TFV. The stock of target cells with or without drug was then split into wells at 10^6 cells well^{-1} and incubated for 4 h. After target cell incubation, HLA-A2-negative donor PBMCs were washed, counted, diluted to 10^6 cells ml^{-1} and added to target cells at an approximately 1:10 donor:target ratio as follows. For cell-free infection, each well received 100 μl mock-infected HLA-A2-negative donor PBMCs and 150 μl (250 ng) cell-free NL4-3. For co-culture infection, each well received 100 μl infected HLA-A2-negative donor PBMCs and 150 μl growth medium. One day after target-cell infection, cell aggregates were broken up by repeated pipetting, and cells split 1:2 with fresh growth medium containing the corresponding drug concentration. Two days after target-cell infection, the number of infected target cells was determined: cells were stained with PE-conjugated anti-HLA-A2 antibody (BD Biosciences or Biolegend), fixed and permeabilized (Cytofix/cytoperm kit, BD Biosciences), then stained with intracellular FITC-conjugated anti-HIV p24 antibody (clone KC57, Coulter Corporation) according to the Cytofix/cytoperm kit protocol. The fraction of infected target cells was quantified by FACS as HLA-A2, p24 double-positive cells. We observed that PBMCs were infected best when fresh, and use of previously frozen material or cells whose processing was delayed substantially reduced both cell-free and co-culture infections.

Comparison of co-culture and cell-free infections using Rev-CEM cells. We infected Rev-CEM target cells either by co-culture with MT-4mCherry donor cells or cell-free virus. MT-4mCherry donor cells at 4×10^5 cells ml^{-1} were infected with 300 ng ml^{-1} p24 NL4-3, or mock infected with the same volume of growth medium. Donor cells were then incubated for 3 days. Two days after donor-cell infection and one day before target-cell infection, Rev-CEM target cells at 8×10^5 cells ml^{-1} were treated with no drug, TFV, or EFV. The stock of target cells with or without drug was then split into wells at 1.6×10^6 cells well^{-1} and incubated for 24 h. Three days after donor-cell infection, MT-4mCherry donor cells were washed, counted, diluted to 3×10^5 cells ml^{-1} and added at an approximately 1:100 donor:target ratio as follows. For cell-free infection, each well received 100 μl mock-infected MT-4mCherry donor cells and 600 μl (1 μg) cell-free NL4-3. For co-culture infection, each well received 100 μl infected MT-4mCherry donor cells and 600 μl growth medium. One day after target-cell infection, cell aggregates were broken up by repeated pipetting, and cells split 1:2 with fresh growth medium containing the corresponding drug concentration. Two days after target-cell infection, the number of infected target cells were quantified by FACS by mCherry and GFP fluorescence. Infected target cells were gated as positive for GFP, and negative for mCherry, thereby excluding uninfected Rev-CEM cells (GFP negative), MT-4mCherry cells (GFP negative, mCherry positive), and fusions between MT-4mCherry and Rev-CEM cells (GFP positive, mCherry positive). The fraction of MT-4mCherry donors was 1% on day 0 for both mock-infected and infected cells, but decreased for infected MT-4mCherry cells by the end of the target-cell infection, probably owing to the cytotoxicity of infection. To ensure that the low numbers of infected target cells gave repeatable results, we averaged consecutive independent inter-day experiments.

Infection growth rate. To initiate infection, Rev-CEM cells at 4×10^5 cells ml^{-1} were infected with 300 ng ml^{-1} NL4-3 in the absence of drugs and incubated for three days. Two days post-infection, uninfected Rev-CEM cells at 8×10^5 cells ml^{-1} were pre-treated with no drug, $100\,\mu M$ TFV, or a combination of EFV, TFV and FTC at their clinical maximum plasma concentrations (C_{max}: $10\,\mu M$ EFV, $2\,\mu M$ TFV and $10\,\mu M$ FTC). Three days after the initial infection with cell-free virus, the infected Rev-CEM cells were washed and added to a final fraction of 0.2% GFP-expressing donor cells to the uninfected Rev-CEM cells incubated with no drug, TFV or C_{max}. On each day after infected donor addition, cell aggregates were broken up by gentle repeated pipetting and cells split. Infection conditions were calibrated so that the number of uninfected target cells would not be limiting and infection would not interfere with proliferation of uninfected cells. Infection was therefore kept below ~0.5% GFP-expressing infected Rev-CEM cells. The daily cell dilution was calibrated to keep this steady state of infected cells: the sample with no drug was split 1:10 or 1:20 into fresh Rev-CEM cells in a new well. For $100\,\mu M$ TFV or C_{max} drug concentrations, infected cells were split 0.6:1 with drug-containing medium into a new well. Cells remaining after cell split were used to quantify the fraction of infected cells by FACS (5×10^5 collected per sample). The fold change in infected cells on each day was calculated as $N_k D_k/N_0$, where N_k is the fraction of infected cells on day k, D_k is the total dilution factor up to day k and N_0 is the fraction of infected cells on day 1. The drug effect on a single round of cell-free infection for $100\,\mu M$ TFV or C_{max} was measured at the same time as the infection growth rate to prevent differences in drug stock batch or cells.

Stochastic simulation of the number of infected cells and mutations. The purpose of the simulation was to determine the sum of total infected cells, newly infected cells, and mutations at each virus cycle (measured as 1.7 days (Supplementary Fig. 2)) from overlapping infection chains. A new infection chain was initiated each virus cycle with an input of one infected cell. The number of infected cells in cycle $k+1$ generated by infected cell j in cycle k was an integer $I_{k+1}^j = x_1 + x_2$, where x_1 was a random number from a Poisson distribution with an average μ_1 defined by the measured infected-cell half-life $\mu_1 = 2^{-\tau/t1/2} = 0.46$ (Supplementary Fig. 13 and Supplementary Theory, section 3), and x_2 was a random number from a Poisson distribution with an average μ_2 defined by $\mu_2 = R_{C_{max}} - \mu_1 = 0.49$ (Supplementary Theory, section 3). Given an outcome of N infected cells in cycle k, the number of total infected cells in virus cycle $k+1$ in the infection chain was $\sum_1^N (x_1^j + x_2^j)$, of which the number of newly infected cells was $\sum_1^N x_2^j$. A new infection chain from an input of one infected cell was generated every virus cycle. Therefore, infection chains overlapped, and the total output number of infected cells in virus cycle $k+1$ was a sum of infected cells at that virus cycle from all M infection chains: $\sum_1^M \sum_1^N (x_1^j + x_2^j)$. The number of newly infected cells was $\sum_1^M \sum_1^N x_2^j$. If a new infection occurred, the probability of mutation occurring at any one of the 10^4 nucleotides of the HIV genome was $1 - (1 - 3.4 \times 10^{-5})^{10^4} = 0.29$, where 3.4×10^{-5} is the per-base probability of mutation for the HIV reverse transcriptase, and $(1 - 3.4 \times 10^{-5})^{10^4}$ is the probability that no mutations occur during a single reverse transcription event. As a simplifying assumption, no fitness benefit or cost was assigned to individual mutations. Therefore, $R_{C_{max}}$ did not change during the course of the simulation. Surviving cells carried over their mutations to the next generation, and newly infected cells carried over mutations from the infected donor cells, in addition to any mutations generated during the infection process.

LETTER

Intravenous delivery of a multi-mechanistic cancer-targeted oncolytic poxvirus in humans

Caroline J. Breitbach[1], James Burke[2]†, Derek Jonker[3,4], Joe Stephenson[5], Andrew R. Haas[6], Laura Q. M. Chow[3,4], Jorge Nieva[2], Tae-Ho Hwang[7], Anne Moon[1], Richard Patt[8], Adina Pelusio[1], Fabrice Le Boeuf[3], Joe Burns[3,4], Laura Evgin[3,4], Naomi De Silva[3,4], Sara Cvancic[3,4], Terri Robertson[1], Ji-Eun Je[7], Yeon-Sook Lee[7], Kelley Parato[3], Jean-Simon Diallo[3], Aaron Fenster[9], Manijeh Daneshmand[3,4], John C. Bell[3,4]* & David H. Kirn[1]*

The efficacy and safety of biological molecules in cancer therapy, such as peptides and small interfering RNAs (siRNAs), could be markedly increased if high concentrations could be achieved and amplified selectively in tumour tissues versus normal tissues after intravenous administration. This has not been achievable so far in humans. We hypothesized that a poxvirus, which evolved for blood-borne systemic spread in mammals, could be engineered for cancer-selective replication and used as a vehicle for the intravenous delivery and expression of transgenes in tumours. JX-594 is an oncolytic poxvirus engineered for replication, transgene expression and amplification in cancer cells harbouring activation of the epidermal growth factor receptor (EGFR)/Ras pathway, followed by cell lysis and anticancer immunity[1]. Here we show in a clinical trial that JX-594 selectively infects, replicates and expresses transgene products in cancer tissue after intravenous infusion, in a dose-related fashion. Normal tissues were not affected clinically. This platform technology opens up the possibility of multifunctional products that selectively express high concentrations of several complementary therapeutic and imaging molecules in metastatic solid tumours in humans.

Despite recent advances in cancer treatment, truly innovative approaches are required to move beyond the modest benefits achieved to date. One novel strategy is the use of replication-competent oncolytic viruses that selectively infect tumours[2,3]. Vaccinia and other poxviruses have several biological properties that make them ideally suited for intravenous delivery and subsequent amplification of transgenes within tumours[1]. First, vaccinia has evolved mechanisms for intravenous stability and spread to distant tissues, including resistance to antibody- and complement-mediated neutralization in the blood[4,5]. Second, vaccinia has also evolved for rapid and motile spread within tissues[6,7]. Because of their relatively large size, vaccinia virions may preferentially deposit in tumours, where neovasculature has increased permeability. Finally, the replication of vaccinia virus is dependent on EGFR/Ras pathway signalling[8,9] which is commonly activated in epithelial cancers[10].

JX-594 is a Wyeth strain vaccinia-vaccine-derived oncolytic virus engineered for viral thymidine kinase gene inactivation, and expression of transgenes encoding human granulocyte-macrophage colony stimulating factor (GM-CSF, encoded by *CSF2*) and β-galactosidase (β-gal, encoded by *lacZ*) under control of the synthetic early–late and p7.5 promoters, respectively[11,12]. Selective replication of the virus in cancer cells is driven by cellular EGFR/Ras pathway signalling, thymidine kinase elevation and type-1 interferon resistance[1,9,13]. In a Phase 1 trial of intratumoural injection into liver tumours, JX-594 was well-tolerated and associated with replication, expression of biologically active hGM-CSF and tumour destruction[14].

The clinical trial described herein was designed to test whether JX-594 could infect metastatic tumours after intravenous infusion in patients. First, we assessed JX-594 selectivity for tumour tissue after *ex vivo* infection of paired explants of viable tumour and adjacent normal tissue obtained from patients undergoing surgery. Within 24 h of exposure, JX-594 was able to infect tumour tissue selectively in seven of ten samples: most tumours had high-intensity staining whereas normal tissues did not (Fig. 1). Peripheral blood mononuclear cells were also highly resistant to infection (data not shown).

We subsequently performed a Phase 1 dose-escalation trial of a single intravenous infusion of JX-594 in 23 patients with advanced, treatment-refractory solid tumours (Table 1). Patients were treated in one of six dose cohorts (1×10^5 to 3×10^7 plaque-forming units (p.f.u.) kg^{-1}). JX-594 delivery, gene expression and replication in solid tumours were assessed. Safety (including determination of the maximum tolerated dose (MTD) or maximum feasible dose (MFD)), pharmacokinetics and antitumour activity were also evaluated.

For pharmacokinetic analyses, quantitative PCR (qPCR) was used to measure genome concentrations in blood during the 1 h of intravenous
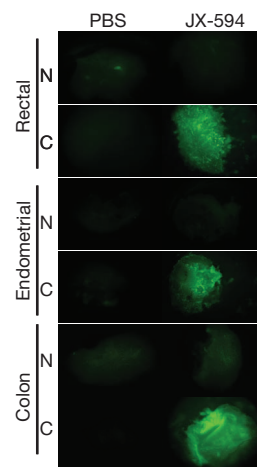
Figure 1 | *Ex vivo* infection of explants of tumour and normal tissue from patients reveals tumour-selective JX-594 gene expression. JX-594 expressing green fluorescent protein (GFP) (JX-594-GFP$^+$/β-gal$^-$) was used to infect primary live-tissue specimens from cancer patients undergoing surgical resection. Matched tumour and adjacent normal tissues were infected overnight to assess the selectivity of transgene expression and replication, or were treated with PBS as a negative control. GFP expression from JX-594-GFP$^+$/β-gal$^-$-infected cells was assessed using a fluorescence microscope. N = normal tissue, C = cancer tissue.

[1]Jennerex Inc., 450 Sansome Street, 16th floor, San Francisco, California 94111, USA. [2]Department of Hematology/Oncology, 801 North 29th Street, Billings Clinic, Billings, Montana 59101, USA. [3]Ottawa Hospital Research Institute, 501 Smyth Road, Ottawa, Ontario K1H 8L6, Canada. [4]University of Ottawa, 75 Laurier Avenue East, Ottawa, Ontario K1N 6N5, Canada. [5]Cancer Centers of the Carolinas, 3 Butternut Drive, Greenville, South Carolina 29605, USA. [6]University of Pennsylvania Medical Center, 3400 Spruce Street, Philadelphia, Pennsylvania 19104, USA. [7]Pusan National University, Jangjeon-dong GeumJeong-gu, Busan 609-735, South Korea. [8]RadMD, 712 Hyde Park, Doylestown, Pennsylvania 18901, USA. [9]Robarts Research Institute, 100 Perth Drive, P.O. Box 5015, London, Ontario N6A 5K8, Canada. †Present address: Jennerex Inc., 450 Sansome Street, 16th floor, San Francisco, California 94111, USA.
*These authors contributed equally to this work.

Table 1 | Overview of patient characteristics, JX-594 delivery to tumours and antitumour activity

	Patient characteristics								JX-594 delivery, replication			Anti-tumour activity		
Pt	Tumour type	Dose (p.f.u. kg^{-1})	Age (yrs)	Gender	Previous lines of therapy	Metastatic tumour sites	Sum LD (cm)	BL NAb	Anti-β-gal (day 29)	Biopsy (day 8–10) IHC	Biopsy (day 8–10) PCR	DC* (RECIST)	DC Duration	Mod. Choi
01	Lung (NSCLC)	1×10^5	64	F	2	6	16.2	NEG	NEG	NEG	NEG	No (PD)	NA	NR
02	Colorectal	1×10^5	72	M	5	3	14.7	POS	NEG	NEG	NEG	No (PD)	NA	NR
03	Lung (NSCLC)	1×10^5	59	F	5	4	3.1	NEG	NEG	NEG	NEG	Yes (SD)	>10 weeks	NR
04	Colorectal	1×10^6	80	F	0	6	22.2	NEG	NEG	NEG	NEG	No (PD)	NA	NR
05	Melanoma	1×10^6	76	M	1	9	26.5	NEG	NEG	NEG	NEG	No (PD)	NA	ND
06	Thyroid	1×10^6	66	M	2	1	3.2	POS	NEG	NEG	NEG	Yes (SD)	>4 weeks	NR
07	Lung (NSCLC)	3×10^6	67	F	4	6	10	POS	NEG	NEG	NEG	Yes (SD)	>4 weeks	NR
08	Pancreatic	3×10^6	75	F	3	4	7.1	NEG	POS	NEG	NEG	No (PD)	NA	R
09	Melanoma	3×10^6	52	M	3	5	13	POS	NEG	NEG	NEG	Yes (SD)	4 weeks	R
10	Colorectal	3×10^6	64	M	2	1	5.4	NEG	NEG	NEG	NEG	Yes (SD)	>10 weeks	NR
11	Ovarian	1×10^7	60	F	5	4	14.4	NEG	NEG	NEG	NEG	Yes (SD)	>4 weeks	NR
12	Melanoma	1×10^7	60	M	4	3	37.1	NEG	ND	NEG	NEG	ND	ND	ND
13	Lung (NSCLC)	1×10^7	67	F	3	6	3.4	POS	NEG	NEG	NEG	Yes (SD)	10 weeks	R
15	Gastric	1×10^7	61	F	5	5	11	NEG	POS	NEG	NEG	No (PD)	NA	R
14	Melanoma	3×10^7	79	F	3	6	15.1	NEG	POS	NEG	NEG	Yes (SD)	>4 weeks	ND
16	Leiomyosarcoma	3×10^7	55	F	3	6	17	NEG	POS	POS	NEG	Yes (SD)	>16 weeks	NR
17	Lung (NSCLC)	3×10^7	43	M	3	4	13.5	NEG	POS	POS	POS	No (PD)	NA	ND
18	Ovarian	1.5×10^7†	61	F	7	5	13	NEG	POS	POS	POS	Yes (SD)	>16 weeks	ND
20	Colorectal	1.5×10^7	67	F	3	8	22	NEG	POS	POS	POS	No (PD)	NA	R
21	Colorectal	1.5×10^7	57	F	2	7	17.8	NEG	POS	POS	POS	Yes (SD)	4 weeks	ND
22	Mesothelioma	1.5×10^7	68	F	2	4	12	NEG	POS	NEG	POS	Yes (PR‡)	>10 weeks	ND
23	Colorectal	1.5×10^7	68	M	3	3	7.8	POS	POS	POS	NEG	Yes (SD)	>4 weeks	ND

Patient 19 was not evaluable. Anti-β-gal, antibody development to the product of the β-galactosidase marker transgene between baseline and day 29; BL, baseline; DC, disease control; LD, longest diameter of tumour; Mod. Choi, modified Choi response; NA, not applicable; NAb, neutralizing antibody to JX-594 (due to previous vaccinia vaccination); ND, not determined; NEG, negative; NR, no response; NSCLC, non-small-cell lung cancer; PD, progressive disease; POS, positive; PR, partial response; Pt, Patient number; R, response; SD, stable disease.
* Stable disease or partial response by RECIST criteria at week 4 and/or 10.
† Approximate dosage based on 66.7 kg weight (patients were treated at 1×10^9 p.f.u. (fixed dose)).
‡ Partial response determined according to modified RECIST for mesothelioma[17].

a, Acute pharmacokinetics of JX-594 genomes (qPCR) after a single intravenous infusion, plotted by dose cohort.

- Cohort 1 (1×10^5 p.f.u. kg^{-1}) ($n = 3$)
- Cohort 2 (1×10^6 p.f.u. kg^{-1}) ($n = 3$)
- Cohort 3 (3×10^6 p.f.u. kg^{-1}) ($n = 4$)
- Cohort 4 (1×10^7 p.f.u. kg^{-1}) ($n = 4$)
- Expansion (1.5×10^7 p.f.u. kg^{-1}) ($n = 6$)
- Cohort 5 (3×10^7 p.f.u. kg^{-1}) ($n = 3$)
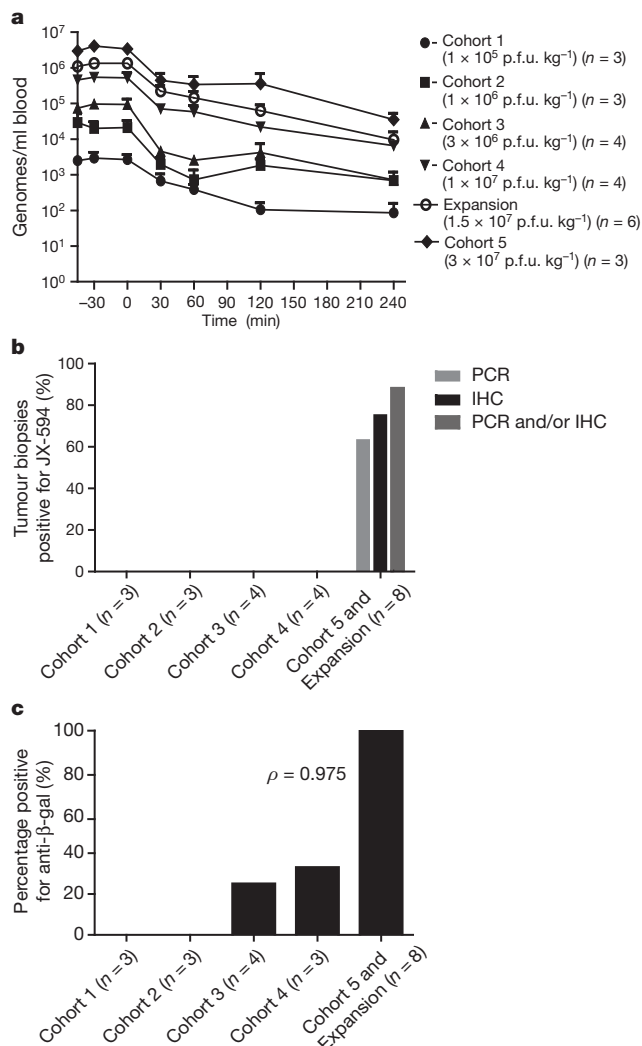
infusion and 4 h afterwards. The highest concentrations were detected during infusion. At doses $\geq 10^7$ p.f.u. kg^{-1} ($\geq 5 \times 10^8$ p.f.u. dose^{-1}), genomes were still detectable in blood at 4 h (Fig. 2a). Peak concentrations and area-under-the-curve were both dose-related. JX-594 infectious units were also detected during intravenous infusion and at 4 h in high-dose patients (data not shown).

JX-594 was generally well tolerated. Dose escalation proceeded without dose-limiting toxicities. Therefore, more than 3×10^7 p.f.u. kg^{-1} (approximately 2×10^9 p.f.u. dose^{-1}) was the MFD; a fixed dose of 1×10^9 p.f.u. was administered to patients in the expansion cohort ($n = 5$ evaluable patients). The most common treatment-related adverse events were grade 1–2 flu-like symptoms lasting up to 24 h: fever (78%), chills (56%), fatigue, headache, nausea, hypotension (22% each), vomiting (17%), tachycardia, hypertension, anorexia and myalgia (13% each). A single grade-1 skin pustule was noted in each of two patients one week after infusion and resolved without sequelae. Levels of interferon (IFN)-γ, tumour necrosis factor (TNF)-α and interleukin (IL)-6 increased acutely in a dose-dependent manner (peak, 8 h; resolution, day 4); levels of IL-10 increased on days 4–8. In contrast, IL-1 did not change significantly and IL-4 decreased transiently (data not shown). Neutralizing antibodies to vaccinia were detectable in six patients at baseline, and titres increased by day 15 in all high-dose patients. No correlation was demonstrated between antibody titres (baseline or induced) and JX-594 replication, safety or antitumour activity. Shedding to the environment was assessed (Supplementary Discussion).

Cancer-selective and dose-related JX-594 delivery and replication in tumours were demonstrated in biopsies obtained 8–10 days after infusion (Fig. 2b and Table 1). Of patients treated at $\geq 1.5 \times 10^7$ p.f.u. kg^{-1}

Figure 2 | JX-594 is selectively delivered to, and amplified within, tumours after intravenous infusion. a, Acute pharmacokinetics of JX-594 genomes (qPCR) after a single intravenous infusion, plotted by dose cohort. Error bars are s.e.m. **b**, Dose-dependent delivery of JX-594, as demonstrated by PCR and/ or immunohistochemical (IHC) analysis of tumour biopsies collected 8–10 days after treatment. **c**, Dose-dependent induction of antibodies to β-galactosidase in patients evaluable for this endpoint. ρ = Spearman's rank correlation coefficient. The number of patients evaluable for each group (n) is indicated.

($\geq 10^9$ p.f.u. dose^{-1}) and evaluable for biopsy analysis, 87% showed JX-594 positivity by qPCR and/or immunohistochemistry, whereas JX-594 was not detected in biopsies collected from subjects treated at lower doses. Of note, delivery and replication were demonstrated in a patient with baseline antibodies to JX-594/vaccinia (Supplementary Fig. 1). Infection resulted in granular cytoplasmic staining by immunohistochemistry, indicative of virus replication ('virus factories') (Fig. 3a, d and Supplementary Fig. 2). Diffuse infection and interspersed necrosis of malignant glandular structures were also evident. Adjacent and intermixed normal tissues were negative for replication by immunohistochemistry (Fig. 3i). In immediately adjacent normal squamous epithelium, low-level diffuse staining without cytopathogenicity indicated uptake without replication. Staining was absent in negative-control biopsy samples (pretreatment tumour from the same patient, Fig. 3c, f; no primary antibody, Fig. 3b, e). Three-dimensional visualization of JX-594 infection in a tumour gland revealed diffuse infection and spread (Fig. 3g, h and Supplementary Movie 1).

Immunohistochemical staining for β-galactosidase protein, the product of the *lacZ* transgene, confirmed expression in JX-594-infected tumour cells (Fig. 3j–l). In addition, the induction of antibodies to β-galactosidase was a surrogate marker for JX-594 replication and transgene expression (β-galactosidase is not present in the product; high-level expression requires replication). Induction of anti-β-galactosidase antibody was dose-related, occurring in 100% of high-dose patients (Spearman's rank correlation coefficient ρ = 0.975; P = 0.005) (Fig. 2c). Tumour-biopsy positivity correlated strongly with β-galactosidase antibody induction (Table 1). Finally, expression of the second JX-594 transgene (*GM-CSF*) was assessed. In a previous clinical

trial, we demonstrated that high-level replication could result in detectable GM-CSF in blood when other inflammatory cytokines had returned to baseline (days 4–15)[14], despite the relatively short half-life of GM-CSF (< 2 h)[15]. Therefore, detection of GM-CSF in blood is indicative of high-level *GM-CSF* expression, and may be a specific but insensitive marker of transgene expression. Three patients had significant increases from baseline in plasma GM-CSF concentrations (days 4–15; all other inflammatory cytokines returned to baseline in < 24 h). All three had evidence of β-galactosidase expression, tumour infection and/or antitumour activity (modified Choi and/or fluoro-deoxy-D-glucose positron emission tomography (FDG-PET) criteria). GM-CSF-protein-responsive subsets of white blood cells (neutrophils, eosinophils and monocytes) peaked on days 4–15 (Supplementary Fig. 3).

Dose-related antitumour activity was demonstrated (modified Choi[16] or response evaluation criteria in solid tumours (RECIST)[17] criteria), and correlated with delivery and replication of JX-594 (Table 1 and Supplementary Fig. 4). Furthermore, new tumour outgrowth in the time after treatment was less frequent in patients treated with high doses than low doses (Spearman's rank correlation coefficient ρ = −0.872; P = 0.05), indicating suppression of microscopic tumour foci. Two out of five high-dose patients had antitumour activity by FDG-PET (>25% decrease in standardized uptake value).

Here we report the first reproducible dose-related delivery, replication and transgene expression from a viral vector or oncolytic virus in metastatic solid tumours in humans after intravenous administration. Engineered oncolytic poxviruses such as JX-594 can express several complementary therapeutic proteins[18] and/or siRNAs[19] in metastatic
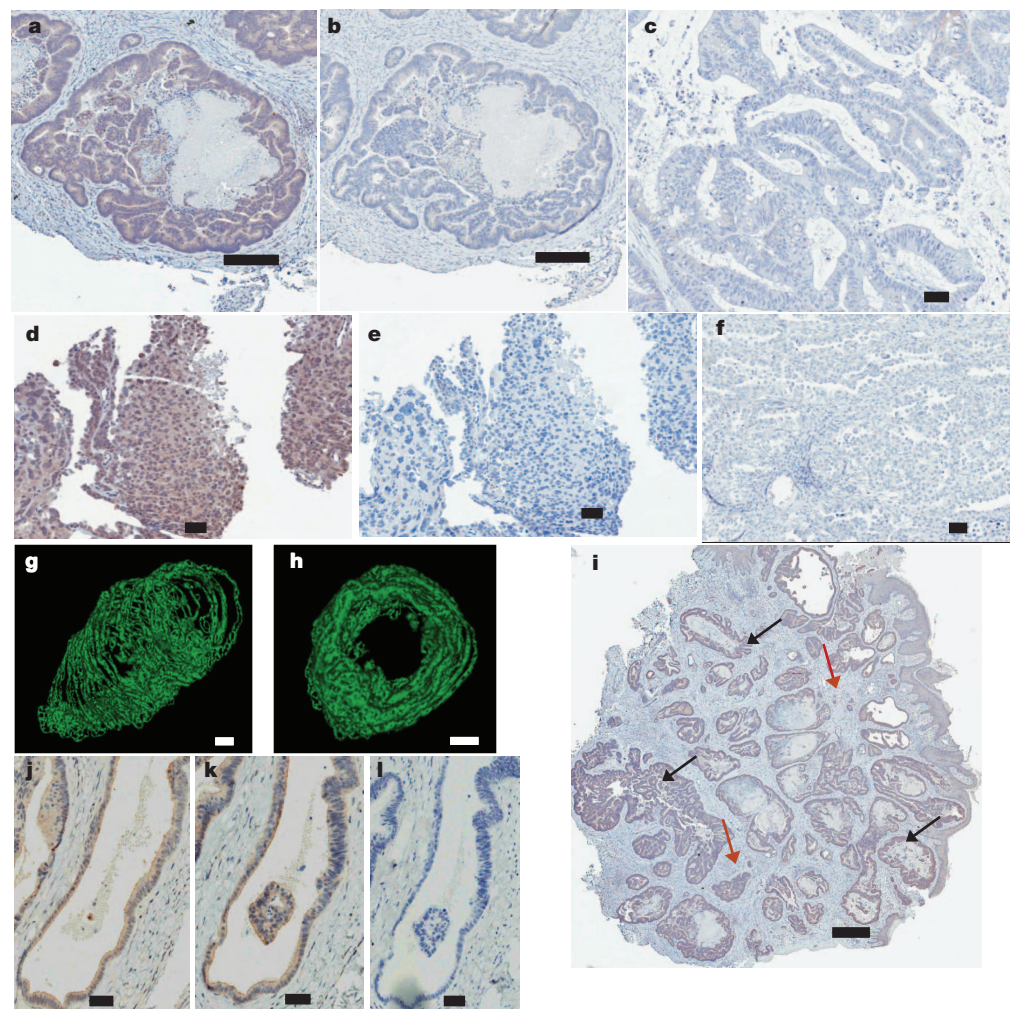
Figure 3 | Immunohistochemical staining reveals JX-594 infection and β-galactosidase expression in tumours. **a**, Immunohistochemistry for vaccinia (patient 20, 10 days after treatment). Scale bar, 200 µm. **b**, Immunohistochemistry, no primary antibody. Scale bar, 200 µm. **c**, Immunohistochemistry for vaccinia in pre-treatment biopsy. Scale bar, 50 µm. **d–f**, As in **a–c** for patient 18, biopsy at 8 days after treatment. Scale bars, 50 µm. **g**, **h**, Three-dimensional reconstruction of vaccinia (green) throughout tumour in patient 20. Scale bars 200 µm. **i**, Immunohistochemistry for vaccinia at low magnification (patient 20). Scale bar, 500 µm. Black arrows indicate tumour; red arrows indicate normal tissue. **j**, Immunohistochemistry for β-galactosidase (patient 20). Scale bar, 50 µm. **k**, Immunohistochemistry for vaccinia. Scale bar, 50 µm. **l**, Negative control. Scale bar, 50 µm. Linear adjustment to brightness and contrast was applied to **j–l**.

tumours systemically and in a cancer-selective fashion; therapeutic concentrations in tumour tissues should therefore be markedly higher than in normal tissues. Incorporation of transgenes encoding marker proteins can facilitate the monitoring of product replication and transgene expression through blood[20] or radiographic assessments[21,22]. Repeated intravenous dosing with JX-594 is currently being assessed and related oncolytic poxviruses should also be tested[23,24]. Although anti-viral immunity may in theory decrease the efficiency of delivery, all patients on this trial had a history of vaccination with live vaccinia virus as children, and delivery was demonstrated in a patient with neutralizing antibodies present at baseline. In preclinical murine tumour models, intravenous vaccinia delivery and efficacy was feasible despite high-titre antibodies[25]. Repeated intravenous delivery may be feasible because of the unique biology of vaccinia, including its ability to produce 'stealth' particles (extracellular enveloped virus) that can traffic in blood in the presence of neutralizing antibodies and complement[4,5]. In addition, intravenous pharmacological dosing of JX-594 constitutes a route and dose that may transiently saturate native mechanisms of viral clearance. JX-594 and related poxvirus constructs thus represent a novel, systemic, multi-functional cancer-biotherapeutic platform.

METHODS SUMMARY

Patients. Twenty-three patients with treatment-refractory, histologically confirmed, advanced, metastatic solid tumours were enrolled and received a single intravenous infusion of JX-594 at one of six dose levels. This trial was registered with clinical trials registration number NCT00625456.

All patients gave written informed consent according to guidelines on good clinical practice. Protocol and consent forms were approved by the United States Food and Drug Administration and Health Canada, as well as the Institutional Review and Infection-Control Committees at each hospital. An independent data-safety monitoring board reviewed the clinical safety data from each patient cohort before each of the four dose escalations.

Tumour biopsy analysis. Biopsies (excisional, core-needle or fine-needle aspirate) were obtained from all subjects 8–10 days after treatment, and were formalin-fixed and paraffin-embedded. Sections were subjected to haemotoxylin and eosin staining, immunohistochemical staining for JX-594 proteins, and PCR for JX-594 genomes. Immunohistochemistry used anti-vaccinia polyclonal antibody (Quartett) and a secondary antibody kit (Vectastain, Vector Laboratories). For detection of β-galactosidase by immunohistochemistry, an anti-β-galactosidase polyclonal antibody (Abcam) was used. Negative controls were run without primary antibody and tumours from mice treated with JX-594 were included as positive controls. For PCR, DNA was extracted from 5×10-μm sections using FormaPure kit (Agencourt) and amplified using primers corresponding to the vaccinia E3L gene, TCCGTCGATGTCTACACAGG and ATGTATCCCGCGAAAAATCA, designed using Primer3 software[26], using QuantiTect SYBR Green PCR kit (Qiagen). Stained sections were digitized using the Aperio Scanscope (Aperio) and analysed using ImageScope software. Adobe Photoshop CS software (Adobe) was used to apply linear adjustments to brightness and contrast across all compared stains where indicated.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Kirn, D. H. & Thorne, S. H. Targeted and armed oncolytic poxviruses: a novel multi-mechanistic therapeutic class for cancer. *Nature Rev. Cancer* **9**, 64–71 (2009).
2. Parato, K. A., Senger, D., Forsyth, P. A. & Bell, J. C. Recent progress in the battle between oncolytic viruses and tumours. *Nature Rev. Cancer* **5**, 965–976 (2005).
3. Liu, T. C., Galanis, E. & Kirn, D. Clinical trial results with oncolytic virotherapy: a century of promise, a decade of progress. *Nat. Clin. Pract. Oncol.* **4**, 101–117 (2007).
4. Vanderplasschen, A., Hollinshead, M. & Smith, G. L. Antibodies against vaccinia virus do not neutralize extracellular enveloped virus but prevent virus release from infected cells and comet formation. *J. Gen. Virol.* **78**, 2041–2048 (1997).
5. Vanderplasschen, A., Mathew, E., Hollinshead, M., Sim, R. B. & Smith, G. L. Extracellular enveloped vaccinia virus is resistant to complement because of incorporation of host complement control proteins into its envelope. *Proc. Natl Acad. Sci. USA* **95**, 7544–7549 (1998).
6. Wein, L. M., Wu, J. T. & Kirn, D. H. Validation and analysis of a mathematical model of a replication-competent oncolytic virus for cancer treatment: implications for virus design and delivery. *Cancer Res.* **63**, 1317–1324 (2003).
7. Smith, G. L., Murphy, B. J. & Law, M. Vaccinia virus motility. *Annu. Rev. Microbiol.* **57**, 323–342 (2003).
8. Katsafanas, G. C. & Moss, B. Vaccinia virus intermediate stage transcription is complemented by Ras-GTPase-activating protein SH3 domain-binding protein (G3BP) and cytoplasmic activation/proliferation-associated protein (p137) individually or as a heterodimer. *J. Biol. Chem.* **279**, 52210–52217 (2004).
9. Yang, H. *et al.* Antiviral chemotherapy facilitates control of poxvirus infections through inhibition of cellular signal transduction. *J. Clin. Invest.* **115**, 379–387 (2005).
10. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
11. Mastrangelo, M. J. *et al.* Intratumoral recombinant GM-CSF-encoding virus as gene therapy in patients with cutaneous melanoma. *Cancer Gene Ther.* **6**, 409–422 (1999).
12. Kim, J. H. *et al.* Systemic armed oncolytic and immunologic therapy for cancer with JX-594, a targeted poxvirus expressing GM-CSF. *Mol. Ther.* **14**, 361–370 (2006).
13. Buller, R. M., Smith, G. L., Cremer, K., Notkins, A. L. & Moss, B. Decreased virulence of recombinant vaccinia virus expression vectors is associated with a thymidine kinase-negative phenotype. *Nature* **317**, 813–815 (1985).
14. Park, B. H. *et al.* Use of a targeted oncolytic poxvirus, JX-594, in patients with refractory primary or metastatic liver cancer: a phase I trial. *Lancet Oncol.* **9**, 533–542 (2008).
15. Hovgaard, D., Mortensen, B. T., Schifter, S. & Nissen, N. I. Clinical pharmacokinetic studies of a human haemopoietic growth factor, GM-CSF. *Eur. J. Clin. Invest.* **22**, 45–49 (1992).
16. Choi, H. *et al.* Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. *J. Clin. Oncol.* **25**, 1753–1759 (2007).
17. Byrne, M. J. & Nowak, A. K. Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. *Ann. Oncol.* **15**, 257–260 (2004).
18. Moss, B. Vaccinia virus: a tool for research and vaccine development. *Science* **252**, 1662–1667 (1991).
19. Davis, M. E. *et al.* Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. *Nature* **464**, 1067–1070 (2010).
20. Peng, K. W. *et al.* Intraperitoneal therapy of ovarian cancer using an engineered measles virus. *Cancer Res.* **62**, 4656–4662 (2002).
21. McCart, J. A. *et al.* Oncolytic vaccinia virus expressing the human somatostatin receptor SSTR2: molecular imaging after systemic delivery using 111In-pentetreotide. *Mol. Ther.* **10**, 553–561 (2004).
22. Msaouel, P. *et al.* Noninvasive imaging and radiovirotherapy of prostate cancer using an oncolytic measles virus expressing the sodium iodide symporter. *Mol. Ther.* **17**, 2041–2048 (2009).
23. McCart, J. A. *et al.* Systemic cancer therapy with a tumor-selective vaccinia virus mutant lacking thymidine kinase and vaccinia growth factor genes. *Cancer Res.* **61**, 8751–8757 (2001).
24. Thorne, S. H. *et al.* Rational strain selection and engineering creates a broad-spectrum, systemically effective oncolytic poxvirus, JX-963. *J. Clin. Invest.* **117**, 3350–3358 (2007).
25. Thorne, S. H. *et al.* Targeting localized immune suppression within the tumor through repeat cycles of immune cell-oncolytic virus combination therapy. *Mol. Ther.* **18**, 1698–1705 (2010).
26. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).

METHODS

***Ex vivo* infection of tumour explants.** Primary cancer and normal tissue specimens were obtained from consenting patients who underwent tumour resection. The institutional review board of Ottawa Hospital Research Institute has approved all human studies. A total of ten tumour samples and accompanying adjacent normal tissue samples from the affected organ were assayed for JX-594-GFP$^+$/β-gal$^-$ sensitivity. Samples were received in cell culture medium and processed within 2–16 h. Samples were manually divided using a 15-mm scalpel blade into approximately 10-mm^3 pieces and placed on 12-well plates with minimum essential medium (α-MEM) containing 10% fetal bovine serum (FBS), under sterile techniques. Samples were inoculated with 1×10^7 p.f.u. JX-594 in 100 μl of α-MEM serum-free medium. JX-594 was allowed to adsorb for 2 h at 37 °C, then 1.9 ml of α-MEM supplemented with 10% FBS was added. Infected tissue specimens were incubated in a humidified incubator at 37 °C for 24 h before imaging JX-594-GFP$^+$/β-gal$^-$-driven GFP expression.

Patients. Between 21 July 2008 and 9 February 2010, 23 patients were enrolled. Patients had treatment-refractory, histologically-confirmed, advanced, metastatic solid tumours. At least one tumour mass had to be amenable to biopsy and/or fine-needle aspirate. Patients had adequate haematological function (leukocyte count >3,500 cells per mm^3, CD4 count ≥200 per mm^3, haemoglobin ≥0.1 g ml^{-1}, platelet count ≥100,000 platelets per mm^3) and adequate organ function (including aspartate aminotransferase/alanine aminotransferase <2.5 × upper normal limit, bilirubin ≤1.5 × upper normal limit and serum chemistries within normal limits or Grade 1), a coagulation status of (international normalized ratio (INR) ≤ upper normal limit + 10%), and Karnofsky performance score of ≥70.

Exclusion criteria included known central-nervous-system malignancy, clinically significant and/or rapidly accumulating ascites, pericardial and/or pleural effusions, unstable cardiac disease, increased risk of vaccination complications (exfoliative skin conditions such as eczema or ectopic dermatitis), clinically significant immunodeficiency, anticancer therapy within the preceding four weeks, and pulse oximetry O$_2$ saturation <90% at rest.

All patients gave written informed consent according to good clinical practice guidelines. Protocol and consent forms were approved by the United States Food and Drug Administration and Health Canada, as well as the Institutional Review and Infection-Control committees at each hospital. An independent data-safety monitoring board reviewed the clinical safety data from each patient cohort before each of the four dose escalations.

Manufacturing, product characterization and release testing. Clinical trial material (CTM) lots ($n = 2$) that were used in this study were manufactured according to good manufacturing-practice guidelines. Virus was grown in adherent mammalian cells and purified through sucrose-gradient centrifugation or by tangential flow filtration. *In vitro* and *in vivo* comparability testing demonstrated equivalence of the two lots. Quality control tests on the final product included assays for sterility and endotoxin, DNA, protein, p.f.u. and genome concentration in the CTM; functional assays included potency and GM-CSF production. CTM was formulated in either phosphate-buffered saline with 10% v/v glycerol (pH 7.1), or 30 mM Tris with 10% (w/v) sucrose (pH 7.7). Immediately before the intravenous infusion, JX-594 was diluted in bicarbonate-buffered saline in a total infusion volume of 250 ml.

Treatment. JX-594 was infused in 250 ml bicarbonate-buffered saline over 60 min. Patients in the dose-escalation portion of the trial received one of six dose levels (1×10^5 p.f.u. kg^{-1}, cohort 1; 1×10^6 p.f.u. kg^{-1}, cohort 2; 3×10^6 p.f.u. kg^{-1}, cohort 3; 1×10^7 p.f.u. kg^{-1}, cohort 4; 3×10^7 p.f.u. kg^{-1}, cohort 5; and 1×10^9 p.f.u., expansion cohort) in a group-sequential dose-escalation design (standard 3×3 design; 2–6 patients for each dose cohort) at one of four sites: the Billings Clinic, the Cancer Center of the Carolinas, the Ottawa Hospital Research Institute and the University of Pennsylvania. One additional patient could be treated in each dose cohort once that cohort was cleared for safety. An additional six patients were enrolled as part of an expansion cohort at the approximate midpoint between the cohort-4 and cohort-5 doses, a fixed dose of 1×10^9 p.f.u. (approximately 1.5×10^7 p.f.u. kg^{-1}, depending on patient weight). A starting dose of 1×10^5 p.f.u. kg^{-1} was chosen on the basis of findings from a Good Laboratory Practice toxicology study, and demonstrated safe patient-blood concentrations after intratumoural injection of JX-594 into liver-based tumours[14]. Dose-limiting toxicities were defined as any of the following treatment-related adverse events (through the 4-week evaluation period): 1) any grade-4 toxicity (except isolated grade-4 lymphopenia lasting ≤7 days); 2) grade-3 or -4 hypotension, disseminated intravascular coagulation or allergic reaction/hypersensitivity; 3) grade-3 non-haematological toxicity persisting >7 days (except if toxicity is transaminitis, which may last >7 days if total bilirubin is normal or grade-1, or flu-like symptoms that respond to standard treatments); and 4) grade-3 haematologic toxicity persisting for >7 days (except isolated lymphopenia). The MTD was defined as the dose immediately preceding that for which two or more dose-limiting

toxicities were recorded. If no MTD was defined, the highest dose was defined as the MFD.

Starting with a subset of patients in cohort 4, patients were pre-medicated with acetaminophen (and also treated every 6 h as needed). Patients were advised to hydrate orally for 24 h before treatment (for example, ≥1 litre of solute-containing fluids), and they received hydration through 24 h after treatment (for example, 1–2 litres of fluids, orally or intravenously). Patients were monitored after treatment for 24 h in the hospital and for at least 29 days as outpatients. A physical exam and an interval medical history were performed on each weekly study visit. Safety monitoring included adverse-event monitoring (National Cancer Institute common toxicity criteria, version 3.0) and standard laboratory toxicity grading for haematology, liver and renal function, coagulation studies, serum chemistry and urinalysis.

Pharmacokinetic measurements in blood. JX-594 pharmacokinetics as well as gene expression and replication in solid tumours were assessed. JX-594 genome concentration in blood was measured by qPCR as previously described[14,27]. The pharmacokinetics of JX-594 during and immediately after administration were determined by qPCR. Whole EDTA-blood samples were taken before the start of infusion; 15, 30 and 60 min after the start of infusion; and 30, 60, 120 and 240 min after the end of infusion. Area under the curve (AUC), maximum concentration (C max) and half-life ($t_{1/2}$) were calculated using WinNonLin, version 5.2 (WinNonLin copyright 2010, Pharsight Corporation). The AUC was calculated by the linear trapezoidal method and C max was determined directly by inspection.

Antibody titres to the β-galactosidase marker transgene. Serum samples were obtained at baseline and on days 15 and 29. Human IgG antibodies to β-gal were measured by ELISA. Briefly, plates (NUNC MaxiSorp, Thermo Fisher Scientific) with wells containing β-galactosidase (Sigma) or bicarbonate/carbonate buffer only were incubated overnight at 4 °C and washed with PBS-Tween before incubation with blocking buffer (PBS with 1% bovine serum albumin (BSA), ELISA grade (Sigma)). Diluted serum (1:50, 1:100 or 1:200 in PBS with 0.05% Tween and 1% BSA) was added to β-galactosidase-coated and control wells in duplicate and incubated at 23 °C. Plates were washed and incubated with alkaline-phosphatase-labelled goat anti-human IgG (Abcam) diluted 1:2,000. After washing, the colorimetric substrate alkaline phosphatase yellow (pNPP, Sigma) was added, and NaOH was added to stop colour development after 10 min. Absorbance was read at 405 nm, and absorbance at 630 nm was subtracted. Control-well values were subtracted to account for non-specific binding, and titres values were calculated by comparison to a standard curve of positive sera, arbitrarily assigned a titre of 8,000.

Neutralizing antibody titres to vaccinia virus. This procedure is based on the ability of neutralizing antibodies in patient serum samples to reduce the cytopathic effect caused by live vaccinia virus. Serum samples obtained at baseline and on days 4, 8, 15, 22 and 29 were heat-inactivated, serially diluted in 96-well format (dilution factor 10–3,200,000) and incubated with vaccinia virus for 2 h before transfer of the mixture onto monolayers of A2780 cells. Cell viability was measured 3 days after inoculation by a colorimetric assay based on live-cell-mediated reduction of tetrazolium salt to formazan (Cell Counting Kit-8, Donjindo Laboratories). The neutralizing antibody titre was defined as the reciprocal of the highest dilution of serum that results in ≥50% cell viability.

White blood cell induction. White blood cell count was performed by routine laboratory testing and was included in the safety assessment as defined in the protocol. White blood cell counts were assessed at baseline and on days 4, 8, 15, 22 and 29.

Cytokine measurements. The endogenous cytokines IL-1β, IL-4, IL-6, IL-10, TNF-α and IFN-γ were measured in multiplex in plasma samples obtained at baseline, 3 h and 8 h after dosing, and on days 4 and 8 using a Milliplex kit as directed by the manufacturer (Millipore). GM-CSF concentrations in plasma were determined at baseline, 3 h and 8 h after dosing and on days 4, 8, 15, 22 and 29 using the Quantikine hGM-CSF sandwich ELISA kit as directed by the manufacturer (R&D Systems).

Tumour biopsy analysis. Biopsies (excisional, core-needle, or fine-needle aspirate) were obtained from all subjects 8–10 days after treatment, formalin-fixed and paraffin-embedded. Sections were subjected to haemotoxylin and eosin staining, immunohistochemical staining for JX-594 proteins, and PCR for JX-594 genomes. Immunohistochemistry used anti-vaccinia polyclonal antibody (Quartett) and secondary antibody kit (Vectastain, Vector Laboratories). For detection of β-galactosidase by immunohistochemistry, an anti-β-galactosidase polyclonal antibody (Abcam) was used. Negative controls were run without primary antibody and tumours from mice treated with JX-594 were included as positive controls. For PCR, DNA was extracted from 5×10-μm sections using FormaPure kit (Agencourt) and amplified using primers corresponding to the vaccinia E3L gene, TCCGTCGATGTCTACACAGG and ATGTATCCCGCGAAAAATCA, designed using Primer3 software[26], using QuantiTect SYBR Green PCR kit (Qiagen).

Three-dimensional reconstruction. An excisional biopsy was obtained from patient 20 on day 10 and processed for immunohistochemical detection of vaccinia virus. Serial sections (126) were cut and every other section was stained for virus. Two-dimensional pictures of the sections were then converted into a three-dimensional volume using HTK histology toolkit software (Robarts Imaging Institute). Volume reconstruction was completed using alignment and segmentation-contouring algorithms that oriented each tissue section on top of one another. Each tissue-section image, once oriented, was then converted from two-dimensional pixels into three-dimensional voxels. These three-dimensional stacks were then rendered to generate the reconstructed tumour. Regions of infection were highlighted in green to aid visualization, and in another representation in which the model can be viewed in orthogonal planes, image contrast was adjusted to 50 (Adobe Photoshop CS2, Adobe). Scaling in the three-dimensional model is derived from scaling from two-dimensional images.

Tumour response analysis. Tumour response was assessed by contrast-enhanced computed tomography (CT) imaging on day 29 in all patients, and on week 10 in patients who remained in the study. Maximum tumour diameters and Hounsfield units (density estimate) were obtained at all time points. Patients in the expansion cohort had PET-CT scans done at the same time points; standard uptake values were determined from PET scans. RECIST[28] and modified Choi criteria[16,29] for response were employed in image evaluations. Patient 22 (metastatic mesothelioma) was evaluated by modified RECIST for mesothelioma[17]. The PET response was also determined for PET-evaluable patients.

Equipment and settings. GFP expression from JX-594-GFP$^+$/β-gal$^-$ in human tissue explants was visualized using the Leica M205FA microscope and Leica microsystem LAS AF6000 acquisition software (Leica Microsystem) at ×2–4 magnification. The same exposure time was used across all samples. Images of tumour and normal-tissue samples for each patient were captured on the same day. Images were stored using Adobe Photoshop version 7.0 (Fig. 1).

Immunohistochemistry sections were digitized using the Aperio Scanscope (Aperio) and analysed using ImageScope v10.2.2.2319 software (Aperio). No image adjustments were applied to Fig. 3a–f and i. Adobe Photoshop CS software (Adobe) was used to apply linear adjustments to brightness and contrast across all compared stains where indicated (Fig. 3j–l).

Statistical analysis. The study sample size was set to assess safety issues. The primary objectives were to study the safety and to determine the MTD/MFD of JX-594 after intravenous infusion. Secondary objectives included pharmacokinetics and pharmacodynamics, immune responses (neutralizing antibodies, anti-β-galactosidase antibodies and cytokines) and delivery of JX-594 to solid tumours after intravenous infusion. The likelihood of dose escalation, given variation in true dose-limiting toxicities in the treated population, was calculated as per routine in Phase 1 dose-escalation trials. The expected sample size was 18–24 patients.

The Spearman's correlation coefficient between ranks was used to calculate statistical dependence between antibody induction to β-galactosidase and dose cohort (percentage of patients with antibody induction in cohorts 1–5), as well as appearance of new tumours and dose cohort (percentage of patients with new tumours in cohorts 1–5)[30].

27. Kulesh, D. A. et al. Smallpox and pan-orthopox virus detection by real-time 3′-minor groove binder TaqMan assays on the Roche LightCycler and the Cepheid Smart Cycler platforms. *J. Clin. Microbiol.* **42,** 601–609 (2004).
28. Therasse, P. et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J. Natl. Cancer Inst.* **92,** 205–216 (2000).
29. Choi, H. et al. CT evaluation of the response of gastrointestinal stromal tumors after imatinib mesylate treatment: a quantitative analysis correlated with FDG PET findings. *AJR Am. J. Roentgenol.* **183,** 1619–1628 (2004).
30. Myers, J. L. W. & Arnold, D. *Research Design and Statistical Analysis.* 2nd edn, 505–512 (Laurence Erlbaum, 2003).

LETTER

Modulation of Rab GTPase function by a protein phosphocholine transferase

Shaeri Mukherjee[1]*, Xiaoyun Liu[1]*, Kohei Arasaki[1], Justin McDonough[1], Jorge E. Galán[1] & Craig R. Roy[1]

The intracellular pathogen *Legionella pneumophila* modulates the activity of host GTPases to direct the transport and assembly of the membrane-bound compartment in which it resides[1–6]. *In vitro* studies have indicated that the *Legionella* protein DrrA post-translationally modifies the GTPase Rab1 by a process called AMPylation[7]. Here we used mass spectrometry to investigate post-translational modifications to Rab1 that occur during infection of host cells by *Legionella*. Consistent with *in vitro* studies, DrrA-mediated AMPylation of a conserved tyrosine residue in the switch II region of Rab1 was detected during infection. In addition, a modification to an adjacent serine residue in Rab1 was discovered, which was independent of DrrA. The *Legionella* effector protein AnkX was required for this modification. Biochemical studies determined that AnkX directly mediates the covalent attachment of a phosphocholine moiety to Rab1. This phosphocholine transferase activity used CDP-choline as a substrate and required a conserved histidine residue located in the FIC domain of the AnkX protein. During infection, AnkX modified both Rab1 and Rab35, which explains how this protein modulates membrane transport through both the endocytic and exocytic pathways of the host cell. Thus, phosphocholination of Rab GTPases represents a mechanism by which bacterial FIC-domain-containing proteins can alter host-cell functions.

Legionella pneumophila is an intracellular pathogen that translocates proteins called effectors into the host-cell cytosol using a type IV secretion system called Dot/Icm[8]. The *Legionella* protein DrrA (also known as SidM) is an effector that targets the host GTPase Rab1 (refs 1–3, 5). Initially identified as a Rab1-specific guanine nucleotide exchange factor (GEF), recent studies showed that the amino-terminal region of DrrA has structural similarity to glutamine synthetase adenylyl transferase (GS-ATase) and shares the catalytically important sequence motif G-X_{11}-D-X-D, which enables DrrA to AMPylate the Tyr 77 residue in the class II switch region of Rab1B[7]. To determine if the *in vitro* activity described for DrrA is biologically relevant we examined whether the endogenous DrrA protein mediates Rab1 AMPylation when delivered into host cells during *Legionella* infection.

Cells were infected with a strain of *Legionella* that has a functional Dot/Icm system that delivers effectors into host cells (wild type) or an isogenic Δ*dotA* mutant that has a non-functional Dot/Icm system, and Rab1 protein was analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Fig. 1a). Two different modifications in the switch II region in Rab1 were detected after infection with wild-type *Legionella*. A fragment that corresponded to an AMPylated TITSSYYR peptide (mass to charge ratio, $m/z = 660.5$) was detected. Unexpectedly, a form of this peptide with an unknown moiety of 183 Da ($m/z = 578.5$) was also detected (Fig. 1a). The Δ*dotA* mutant revealed that both modifications required the delivery of effector proteins into host cells during infection. Thus, Rab1 is modified during *Legionella* infection by AMPylation and by a second unknown post-translational mechanism.

Cells were infected with mutant strains of *Legionella* deficient in effectors that could be involved in AMPylation of Rab1. In addition to

the Δ*drrA* mutant, a Δ*ankX* mutant of *Legionella* was examined. The AnkX protein contains a FIC domain, which for other bacterial effectors has been shown to have an enzymatic activity that promotes the AMPylation of small GTPases[9–13]. When microinjected into mammalian cells the AnkX protein disrupts membrane transport in the secretory pathway and interferes with the sorting of transferrin from early endosomes, consistent with AnkX being an effector that disrupts the activities of host membrane transport proteins, potentially by Rab AMPylation[14].

Rab1 AMPylation was not detected in the samples isolated from cells infected with the Δ*drrA* mutant, indicating that DrrA is the primary effector mediating Rab1 AMPylation *in vivo* (Fig. 1a). The unknown modification ($m/z = 578.5$) was detected after infection with the Δ*drrA* mutant, but was not detected after infection with the Δ*ankX* mutant. Thus, the unknown modification to Rab1 that occurs during infection requires AnkX. Defects in Rab1 modifications exhibited by these *Legionella* mutants were complemented upon the introduction of plasmids that restored DrrA and AnkX production (Supplementary Fig. 1a). MS/MS analysis revealed that the unknown 183 Da moiety was attached to Ser 79 of Rab1A, adjacent to the Tyr 80 residue AMPylated by DrrA (Fig. 1b). These residues correspond to Ser 76 and Tyr 77 in Rab1B.

Purified DrrA radiolabelled GST–Rab1 *in vitro* when ^{32}P-α-labelled ATP was used as a substrate, but no labelling was detected using ^{32}P-γ-labelled ATP, validating that DrrA mediates the attachment of AMP to Rab1 (Fig. 1c). The structurally distinct N-terminal region of DrrA was sufficient for AMPylation[3,15–17], and no AMPylation activity was detected for DrrA(340–533) or the DrrA(D110A,D112A) variant having the G-X_{11}-D-X-D adenylyl transferase domain inactivated. The effector AnkX was unable to efficiently AMPylate Rab1, the GTP-locked Rab1(Q70L) variant or the GDP-locked Rab1(S25N) variant, indicating that purified AnkX does not have robust Rab1 AMPylation activity (Fig. 1c).

The nature of the unknown modification to Rab1 requiring AnkX was investigated further. Cells were transfected with a plasmid encoding either AnkX or the variant AnkX(H229A), which has the essential histidine residue in the FIC domain changed to alanine. Roughly 70% of the Rab1 isolated from cells producing AnkX had the 183 Da moiety attached, whereas Rab1 isolated from cells producing the AnkX(H229A) protein was unmodified (Fig. 2a). Thus, AnkX is both necessary and sufficient to promote a novel post-translational modification to Rab1 by a process that requires a functional FIC domain.

For molecules <200 Da, the elemental composition can often be determined from a highly accurate mass measurement[18]. High-resolution MS measurements obtained for the modified Rab1 peptide isolated from cells producing AnkX revealed that the moiety attached to the Ser 79 residue had an accurate mass of 183.0661 Da (Supplementary Fig. 1b). This moiety did not match any known post-translational modifications, but when a metabolite database (http://metlin.scripps.edu) was searched a near perfect match was made to the molecule phosphocholine, which has an exact mass of 183.0660 Da. The protonated moiety attached to the Rab1 peptide was selected and further dissociated by multi-stage MS analysis (MS/MS/MS). The
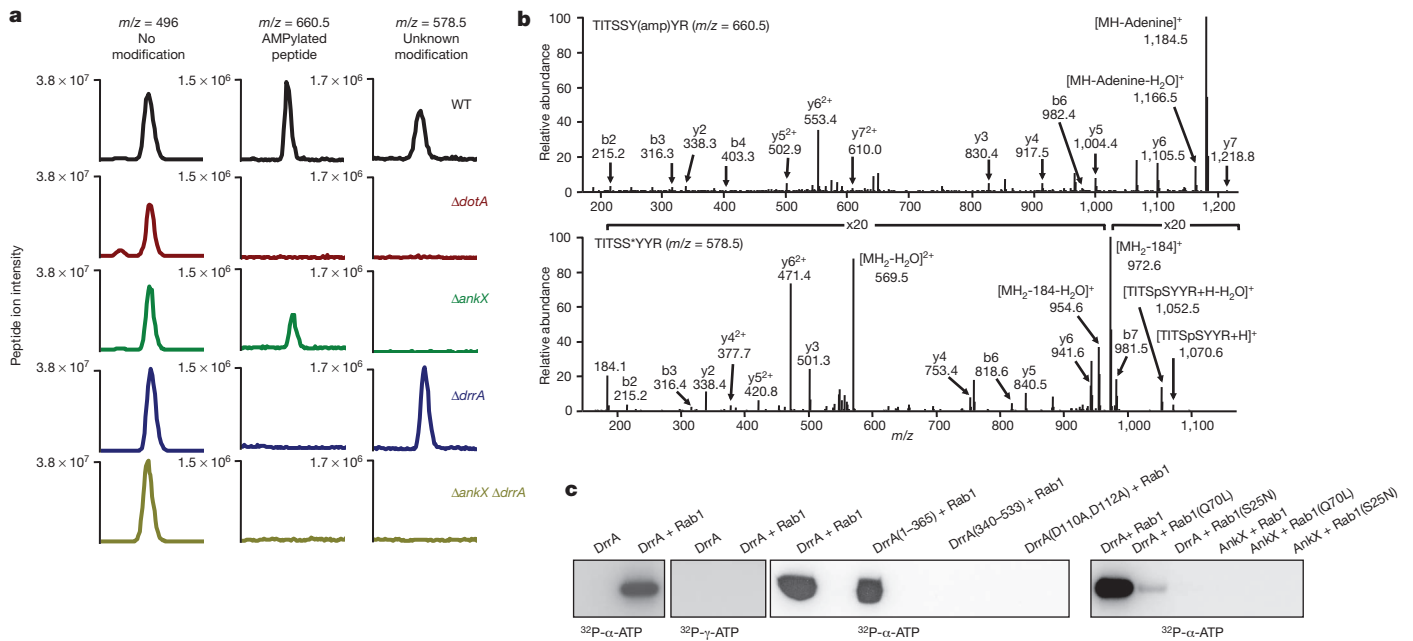
Figure 1 | Legionella infection mediates two different post-translational modifications to Rab1. a, 3×-Flag–Rab1 was isolated from HEK293 FcγRII cells after infection with the indicated *Legionella* strains. LC-MS/MS analysis produced the extracted ion chromatograms. The peak in each graph indicates the amount of Rab1 peptide TITSSYYR with no modifications ($m/z = 496.0$), peptide containing an AMP moiety ($m/z = 660.5$), and peptide with an unknown modification ($m/z = 578.5$). **b**, MS/MS spectra obtained for the AMPylated Rab1 peptide (TITSSY(amp)YR, top) and the Rab1 peptide with the unknown modification (TITSS*YYR, bottom) showing m/z values of their fragments generated were of the sizes predicted for phosphocholine (Fig. 2b) and matched the MS/MS spectrum obtained following dissociation of a phosphocholine standard (Supplementary Fig. 2), suggesting that Rab1 is phosphocholinated by AnkX.

fragments upon collision-induced dissociation. Fragments of the AMPylated peptide (top spectrum) have a mass shift of 329 starting at y3 (y3–y7), indicating a modification site at the first Tyr from the N terminus of the peptide. Fragments of the peptide with the unknown modification have a mass increase of 165 starting at y4 (y4–y6), indicating the second Ser from the N terminus was modified. **c**, Autoradiographs reveal AMPylated Rab1 from *in vitro* reactions containing the recombinant proteins indicated on the top of each panel and the radiolabelled nucleotides indicated below.

If AnkX were functioning directly as a phosphocholine transferase, the host molecule most likely to be used as a substrate in this reaction would be CDP-choline, which is an intermediate used to synthesize phosphatidylcholine[19]. Indeed, phosphocholination of Ser 79 on Rab1 was detected for *in vitro* reactions containing CDP-choline and AnkX, but not in reactions containing DrrA (Fig. 2c) or the AnkX(H229A) protein (Supplementary Fig. 3). Increasing the amount of AnkX in the *in vitro* reaction resulted in higher levels of phosphocholinated Rab1
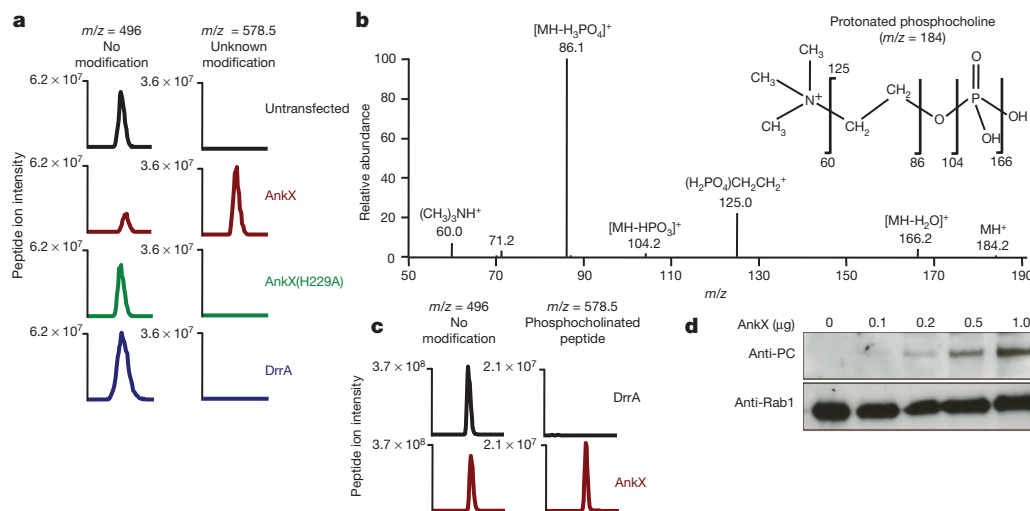
Figure 2 | The Legionella effector AnkX functions as a Rab phosphocholine transferase. a, LC-MS/MS analysis of 3×-Flag–Rab1 isolated from HEK293 cells that were either untransfected or transfected with a plasmid encoding GFP-tagged AnkX, AnkX(H229A) or DrrA. Extracted ion chromatograms indicate the amount of Rab1 peptide TITSSYYR with no modification ($m/z = 496$) and peptide with the unknown modification ($m/z = 578.5$). **b**, MS/MS/MS analysis on the Rab1 peptide TITSSYYR with the unknown modification. The $m/z = 184$ peak corresponding to the protonated moiety attached to the Rab1 peptide was selected and subjected to further dissociation.

Indicated are the fragments identified in the MS/MS/MS spectrum that matched the fragments predicted upon dissociation of the protonated phosphocholine molecule. **c**, The peak in each graph indicates the amount of Rab1 peptide TITSSYYR with no modifications ($m/z = 496$) and phosphocholinated peptide ($m/z = 578.5$) after *in vitro* incubation of Rab1 with either DrrA or AnkX in the presence of CDP-choline. **d**, Immunoblots from *in vitro* reactions that contained Rab1 and the indicated amounts of AnkX. Blots were probed to detect phosphocholinated Rab1 (anti-PC) and total Rab1 (anti-Rab1) in each reaction.
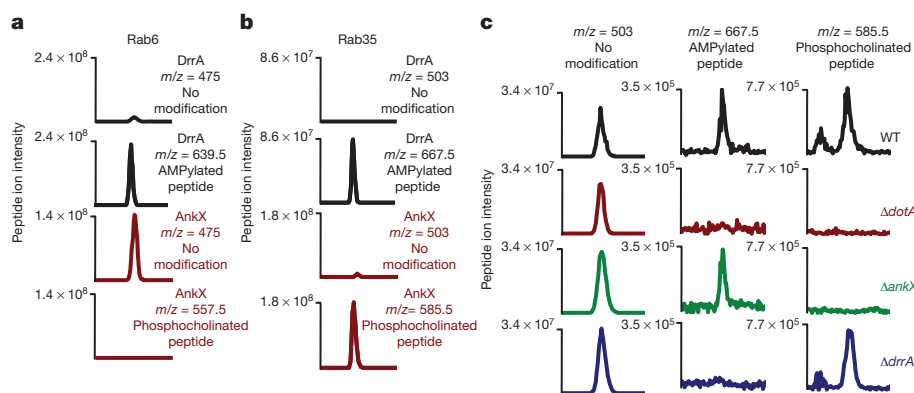
Figure 3 | AnkX and DrrA have overlapping but non-identical Rab specificities. a, Extracted ion chromatograms of 3×-Flag–Rab6 isolated from HEK293 cells producing either DrrA or AnkX. Shown are graphs for the unmodified Rab6 peptide SLIPSYIR ($m/z = 475$) and the AMPylated ($m/z = 639.5$) and phosphocholinated ($m/z = 557.5$) forms of this peptide. **b**, Extracted ion chromatograms of 3×-Flag–Rab35 isolated from HEK293 cells producing either DrrA or AnkX. Shown are graphs for the unmodified Rab35

peptide TITSTYYR ($m/z = 503.0$), the AMPylated peptide ($m/z = 667.5$) and the phosphocholinated peptide ($m/z = 585.5$). **c**, 3×-Flag–Rab35 was isolated from HEK293 FcγRII cells after infection with the indicated *Legionella* strains. The peak in each graph indicates the amount of Rab35 peptide TITSTYYR with no modifications ($m/z = 503$), peptide containing an AMP moiety ($m/z = 667.5$) and phosphocholinated peptide ($m/z = 585.5$). WT, wild type.

being detected by anti-phosphocholine immunoblot analysis (Fig. 2d), validating that Rab1 is phosphocholinated by AnkX.

Phosphocholinated proteins in the size range of Rab GTPases were detected in lysates from cells producing AnkX and were not observed in lysates from cells producing AnkX(H229A) (Supplementary Fig. 4a). Phosphatidylinositol 4-phosphate and phosphatidic acid levels were not affected in cells producing AnkX, suggesting that there is no indirect effect on phospholipid metabolism (Supplementary Fig. 4b). The intracellular pathogen *Coxiella burnetii* translocates a FIC-domain effector called CBU_2078 into host cells[20]. Although proteins reacting with the anti-phosphocholine antibody were found in the size range of small GTPases from cell lysates producing CBU_2078, there was no evidence of Golgi fragmentation or endosome enlargement in these cells (Supplementary Fig. 4a, c). Thus, defects in host membrane transport in AnkX-producing cells probably results from phosphocholination of a specific subset of Rab GTPases, although it cannot be excluded that CDP-choline consumption might augment these effects.

The repertoire of Rab proteins that could be modified by *Legionella* effectors *in vivo* was investigated. Modifications to the Rab5 protein were not detected in cells producing either AnkX or DrrA (Supplementary Fig. 5). DrrA mediated the AMPylation of Rab6 on Tyr 82, but phosphocholination of Rab6 mediated by AnkX was not detected (Fig. 3a and Supplementary Fig. 5). Rab35 is a Rab1 family member that regulates the sorting of cargo from early endosomes, and interfering with Rab35 function results in enlarged early endosomes[21]. Importantly, specific perturbations in Rab35 function result in a cellular phenotype that closely mirrors the defects in endosome morphology observed in cells microinjected with purified AnkX[14]. Phosphocholinated Rab35 was detected in samples isolated from cells producing AnkX, and AMPylated Rab35 was detected from cells producing DrrA (Fig. 3b). During infection, phosphocholination of Rab35 required AnkX and AMPylation of Rab35 required DrrA (Fig. 3c). Thus, AnkX has specificity for Rab1 family members.

To test whether previously described cellular disruptions mediated by AnkX required the FIC-domain-dependent phosphocholine transferase activity (Fig. 4a), cellular phenotypes mediated by AnkX and the AnkX(H229A) mutant were compared. Disruption of the Golgi apparatus and a block in secretion of host alkaline phosphatase into the culture supernatant were observed in cells producing AnkX but not in cells producing AnkX(H229A) (Fig. 4b and Supplementary Fig. 7). Importantly, when the Ser 79 residue in Rab1 was changed to alanine, the variant protein was no longer phosphocholinated by AnkX (Supplementary Fig. 8a); however, Rab1(S79A) interfered with secretion of alkaline phosphatase when produced in cells (Supplementary Fig. 8b).

Thus, AnkX is targeting a residue in Rab1 that is critical for function. There was a significant increase in the number of cells containing enlarged early endosomes in cells producing AnkX compared to cells producing AnkX(H229A) (Fig. 4c and Supplementary Fig. 7), consistent with the function of Rab35 being perturbed by phosphocholination.
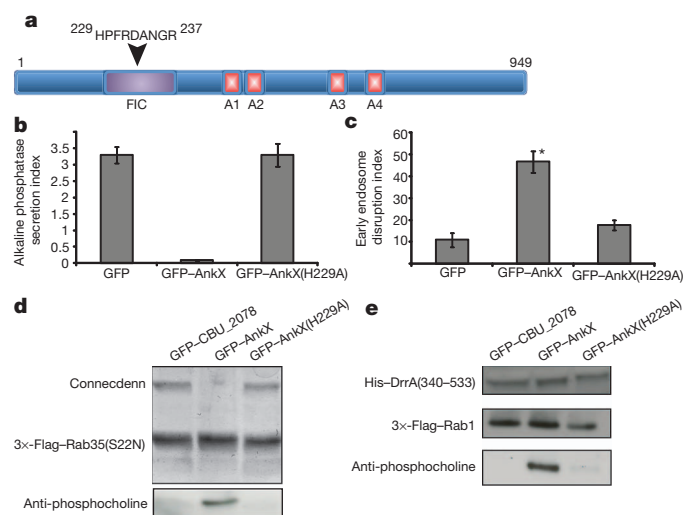
Figure 4 | AnkX-mediated phosphocholination modulates the function of Rab1 and Rab35. a, Schematic representation of the AnkX protein showing the location of the FIC domain and the four predicted ankyrin repeat homology domains (A1–A4). The amino acid sequence in a conserved region of the FIC domain containing the essential His 229 residue is shown. **b**, Secretion of alkaline phosphatase into the culture supernatant was measured for HEK293 cells producing either GFP, GFP–AnkX or GFP–AnkX(H229A) as indicated on the x-axis. Data are the mean ± standard deviation (s.d.) calculated from three independent sample wells. **c**, The disruption of early endosomes was assessed in COS7 cells producing either GFP–AnkX, GFP–AnkX(H229A) or GFP alone after staining for EEA1 (mean ± s.d., $n = 200$, *$P < 0.005$ compared to the GFP alone control). **d**, Binding of recombinant connecdenn to 3×-Flag–Rab35(S22N) isolated from cells producing either GFP–CBU_2078 or GFP–AnkX or GFP–AnkX(H229A) was assessed by co-precipitation. The Coomassie-stained SDS–PAGE gel indicates the amount of connecdenn and 3×-Flag–Rab35(S22N) in each precipitate. The anti-phosphocholine immunoblot indicates that Rab35(S22N) isolated from cells producing GFP–AnkX was phosphocholinated. **e**, Binding of purified His-tagged DrrA(340–533) to 3×-Flag–Rab1A isolated from cells producing either GFP–CBU_2078 or GFP–AnkX or GFP–AnkX(H229A) was assessed by co-precipitation. The immunoblots indicate the amounts of DrrA(340–533) and the levels of phosphocholinated 3×-Flag–Rab1A present in each precipitate.

Because GEF proteins are essential for Rab activation, the effect of phosphocholination on the binding of Rab-specific GEFs was analysed. The eukaryotic connecdenn proteins are the only known GEFs for Rab35 and are required for Rab35 function *in vivo*[21]. A pronounced defect in binding of connecdenn was observed for phosphocholinated Rab35 isolated from cells producing AnkX (Fig. 4d), which would explain why AnkX overproduction mimics the cellular phenotype observed when connecdenn has been silenced in mammalian cells[21]. By contrast, there was no defect in the binding of phosphocholinated Rab1 with the GEF domain of DrrA (Fig. 4e), similar to what has been observed for DrrA interactions with AMPylated Rab1 (ref. 7). Thus, post-translational modifications mediated by the effectors AnkX and DrrA modulate the function of Rab GTPases during infection by tailoring the repertoire of proteins that interact with the modified GTPase.

The characterization of DrrA and AnkX provides an example of *Legionella* having structurally distinct proteins with different biochemical activities that modulate the function of host vesicle transport proteins similarly. This concept of functional redundancy has been postulated but not shown clearly. The differences observed in the *in vivo* specificities shown by these two effectors, however, demonstrate that they are also likely to have roles in modulating Rab protein functions that do not overlap, which could explain why positive selection has led to the emergence of two different pathways to modify Rab protein function through post-translational modification.

The reaction mediated by AnkX has similarities to the AMPylation reaction demonstrated for other FIC domain proteins. Both reactions use a nucleotide-based substrate as the donor molecule that mediates the post-translational modification process (Supplementary Fig. 9). Interestingly, in the AMPylation reaction, hydrolysis of the phosphoanhydrous bond results in protein modification by the 5′-ribonucleotide of the donor substrate, whereas in the phosphocholination reaction the 5′-ribonucleotide is presumably released and the phosphocholine group is transferred to the polypeptide chain.

There are several examples of occasions where post-translational modifications introduced by bacterial toxins or effectors—which were thought to be the exclusive domain of pathogens—were discovered to represent mechanisms used to regulate eukaryotic cell functions. Curiously, the inclusion of a phosphocholine moiety in a protein structure has been indicated previously by studies examining peptides secreted by nematodes and from mammalian cells residing in the placenta[22,23]. Thus, protein phosphocholination may also be used by eukaryotic organisms to modulate cellular functions.

METHODS SUMMARY

MS/MS analysis of Rab GTPases was conducted on immunoprecipitated proteins that were fractionated by SDS–PAGE, digested with trypsin in the gel, and extracted peptides were separated using nano-LC and electrosprayed directly onto a linear ion trap mass spectrometer (LTQ Velos, ThermoElectron) for MS and MS/MS analysis. All biochemical assays were conducted using purified proteins as described in Methods. The antibody TEPC-15 (Sigma) was used to detect phosphocholinated proteins by immunoblot analysis.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Murata, T. *et al.* The *Legionella pneumophila* effector protein DrrA is a Rab1 guanine nucleotide-exchange factor. *Nature Cell Biol.* **8**, 971–977 (2006).
2. Ingmundson, A., Delprato, A., Lambright, D. G. & Roy, C. R. *Legionella pneumophila* proteins that regulate Rab1 membrane cycling. *Nature* **450**, 365–369 (2007).
3. Schoebel, S., Oesterlin, L. K., Blankenfeldt, W., Goody, R. S. & Itzen, A. RabGDI displacement by DrrA from *Legionella* is a consequence of its guanine nucleotide exchange activity. *Mol. Cell* **36**, 1060–1072 (2009).
4. Machner, M. P. & Isberg, R. R. Targeting of host Rab GTPase function by the intravacuolar pathogen *Legionella pneumophila*. *Dev. Cell* **11**, 47–56 (2006).
5. Machner, M. P. & Isberg, R. R. A bifunctional bacterial protein links GDI displacement to Rab1 activation. *Science* **318**, 974–977 (2007).
6. Nagai, H., Kagan, J. C., Zhu, X., Kahn, R. A. & Roy, C. R. A bacterial guanine nucleotide exchange factor activates ARF on *Legionella* phagosomes. *Science* **295**, 679–682 (2002).
7. Muller, M. P. *et al.* The *Legionella* effector protein DrrA AMPylates the membrane traffic regulator Rab1b. *Science* **329**, 946–949 (2010).
8. Ninio, S. & Roy, C. R. Effector proteins translocated by *Legionella pneumophila*: strength in numbers. *Trends Microbiol.* **15**, 372–380 (2007).
9. Yarbrough, M. L. & Orth, K. AMPylation is a new post-translational modiFICation. *Nature Chem. Biol.* **5**, 378–379 (2009).
10. Yarbrough, M. L. *et al.* AMPylation of Rho GTPases by *Vibrio* VopS disrupts effector binding and downstream signaling. *Science* **323**, 269–272 (2009).
11. Kinch, L. N., Yarbrough, M. L., Orth, K. & Grishin, N. V. Fido, a novel AMPylation domain common to Fic, Doc, and AvrB. *PLoS ONE* **4**, e5818 (2009).
12. Roy, C. R. & Mukherjee, S. Bacterial FIC proteins AMP up infection. *Sci. Signal.* **2**, pe14 (2009).
13. Worby, C. A. *et al.* The Fic domain: regulation of cell signaling by adenylylation. *Mol. Cell* **34**, 93–103 (2009).
14. Pan, X., Luhrmann, A., Satoh, A., Laskowski-Arce, M. A. & Roy, C. R. Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* **320**, 1651–1654 (2008).
15. Brombacher, E. *et al.* Rab1 guanine nucleotide exchange factor SidM is a major phosphatidylinositol 4-phosphate-binding effector protein of *Legionella pneumophila*. *J. Biol. Chem.* **284**, 4846–4856 (2009).
16. Zhu, Y. *et al.* Structural mechanism of host Rab1 activation by the bifunctional *Legionella* type IV effector SidM/DrrA. *Proc. Natl Acad. Sci. USA* **107**, 4699–4704 (2010).
17. Suh, H. Y. *et al.* Structural insights into the dual nucleotide exchange and GDI displacement activity of SidM/DrrA. *EMBO J.* **29**, 496–504 (2009).
18. Gross, M. L. Accurate masses for structure confirmation. *J. Am. Soc. Mass Spectrom.* **5**, 57 (1994).
19. Li, Z. & Vance, D. E. Phosphatidylcholine and choline homeostasis. *J. Lipid Res.* **49**, 1187–1194 (2008).
20. Chen, C. *et al.* Large-scale identification and translocation of type IV secretion substrates by *Coxiella burnetii*. *Proc. Natl Acad. Sci. USA* **107**, 21755–21760 (2010).
21. Allaire, P. D. *et al.* The connecDENN DENN domain: a GEF for Rab35 mediating cargo-specific exit from early endosomes. *Mol. Cell* **37**, 370–382 (2010).
22. Lovell, T. M. *et al.* Identification of a novel mammalian post-translational modification, phosphocholine, on placental secretory polypeptides. *J. Mol. Endocrinol.* **39**, 189–198 (2007).
23. Grabitzki, J., Ahrend, M., Schachter, H., Geyer, R. & Lochnit, G. The PCome of *Caenorhabditis elegans* as a prototypic model system for parasitic nematodes: identification of phosphorylcholine-substituted proteins. *Mol. Biochem. Parasitol.* **161**, 101–111 (2008).

Author Contributions S.M., X.L. and K.A. performed research. X.L. conducted the mass spectrometry analysis, K.A. generated HEK293 Fcγ 3×Flag–Rab1 stable cell line, J.M. conducted studies on the CBU_2078 protein, and S.M. conducted all other research. S.M., X.L., J.M. J.E.G. and C.R.R. analysed results and wrote manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.R.R. (craig.roy@yale.edu).

METHODS

Cell culture and transfection. COS7 and HEK293 cells were grown in Dulbecco's modified Eagle medium (DMEM) from Gibco (Carlsbad) containing 10% heat-inactivated fetal bovine serum (FBS; Gibco). Cell lines were cultured at 37 °C in 5% CO_2. For transfection, COS7 or HEK293 cells were added to 12-mm coverslips in 24-well plates at a density of 2.5×10^4 cells per well. Cells were transfected with 0.5 μg of each plasmid. Cells were fixed 18 h after transfection in 4% PFA, permeabilized with 0.05% saponin and processed for immunofluorescence microscopy as described previously[14].

Fluorescence microscopy. Digital images were acquired with a Nikon TE300 microscope using a ×100 1.4 N.A objective lens and a Hamamatsu ORCA-ER camera controlled by IP Lab software.

Alkaline phosphatase secretion assay. Secretion assays were performed as described[24]. Briefly, HEK293 cells were co-transfected with a plasmid encoding secreted alkaline phosphatase and a plasmid encoding either GFP, GFP–AnkX or GFP–AnkX(H229A). After an 18 h incubation period, cells were washed with PBS and fresh medium was added. Cells were incubated for 10 h and then the alkaline phosphatase secretion index was determined by measuring the ratio of alkaline phosphatase protein secreted into the culture medium to the total amount of alkaline phosphatase protein in the assay well. The Tropix PhosphaLight System Kit (Applied BioSystems) was used to measure alkaline phosphatase activity and the Tecan Infinite M1000 plate reader with iControl Software was used to detect chemiluminescence. Data shown are the mean ± s.d. from three independent samples for each condition. Results were validated in two independent assays.

Protein purification and in vitro AMPylation. His-tagged and GST-tagged proteins were purified as described previously[1]. Purified GST–Rab1A (2 μg) was incubated with 0.4 μg of purified effector protein in buffer (20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM $MgCl_2$ and 0.1 mM GTPγS) and incubated for 1 h at 30 °C in the presence of 2 μCi of ^{32}P-α-labelled ATP or ^{32}P-γ-labelled ATP (Perkin Elmer). Labelled proteins were identified by autoradiography following SDS–PAGE. In vitro AMPylation results shown are representative of three independent experiments.

Bacterial strains and plasmids. The *Legionella* strains were grown on charcoal yeast extract plates as described previously[14]. The parental strain (wild type) was *L. pneumophila* serogroup 1 strain Lp01, and the variant strains were all isogenic mutants described previously[1,14,25,26], with the exception of the ΔankX ΔdrrA double mutant, which was generated for this study using allelic exchange to introduce the ΔankX mutation into the ΔdrrA strain as described[14]. For all experiments, *Legionella* were isolated from charcoal yeast extract plates after growth for 2 days at 37 °C. The plasmid pEFGPC2 (Clontech) was used for all GFP fusion constructs, the plasmid pJB1806 was used to produce DrrA and AnkX in *Legionella*, pQE30 (Qiagen) was used for all His-tagged constructs, and pGEX2TK (GE LifeSciences) was used for all GST-tagged constructs. The Rab plasmids were constructed using cDNA encoding human RAB1A, RAB5A, RAB6A, RAB35 and canine RAB1A.

Cell lines and *Legionella* infection. HEK293 FCγRII cells[27] were used to create the HEK293 FCγRII 3×-Flag–Rab1A stable cell line. This cell line was used for assays examining Rab1 modifications during infection of host cells by *Legionella*. For each assay, cells grown to near confluency in two 10-cm dishes were infected with opsonized *Legionella pneumophila* strain LP01 (wild type) or the isogenic mutants at an estimated multiplicity of 100 bacteria to 1 host cell. After incubation for 0.5 h at 37 °C the cells were lysed in buffer containing 20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM $MgCl_2$, 1% Triton X-100, 1 mM PMSF and protease inhibitor cocktail (Roche). Lysates were centrifuged at 17,900*g* and the post-nuclear supernatant was then incubated with Flag-antibody-coated beads (Sigma) at 4 °C for 1 h, washed and the Rab1 protein was eluted using Flag peptide. Samples were then processed for MS/MS analysis. This same approach was used to assay modifications to Rab35 after infection, except that the HEK293 FcγRII cell line was transfected with an expression plasmid producing 3×-Flag–Rab35 and then infected 18 h after transfection.

Golgi and early endosome disruption assay. The disruption of Golgi and early endosomes was assessed in COS7 cells producing GFP–AnkX, GFP–AnkX(H229A), GFP–CBU_2078 or GFP alone. Cells were fixed with 4% PFA 18 h after transfection and stained with mouse monoclonal antibodies specific for either GM130 or EEA1 (BD Transduction Laboratories) at a dilution of 1:200. Golgi and endosome morphology in GFP-positive cells was assessed by fluorescence microscopy. The Golgi was completely fragmented in nearly all the cells producing AnkX, whereas no significant fragmentation was observed compared to the background of untransfected cells in cells producing GFP–AnkX(H229A), GFP–CBU_2078 or GFP alone. The early endosome disruption index represents the percentage of cells producing the indicated GFP protein that showed an enlarged endosome phenotype as determined by EEA1 staining. The data represent the mean ± s.d., from three independent replicates in which 200 cells were counted for each protein. *P* values were computed using Student's unpaired *t*-test. Data shown were validated in two independent experiments.

In vitro phosphocholination assay. GST-tagged Rab1A (5 μg) attached to glutathione agarose was incubated with thrombin to remove the tag. Rab1A was incubated with 1 μg of purified effector protein in buffer (20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM $MgCl_2$, 1 mM ATP) and incubated for 1 h at 30 °C in the presence of 1 mM CDP-choline. Samples were boiled in SDS-loading buffer and the Rab1 protein was excised after SDS–PAGE and analysed by LC-MS/MS analysis. For the immunoblot analysis, the amounts of AnkX in each reaction varied from 0 μg to 1.0 μg, and the amount of modified Rab1 in each reaction was compared by immunoblot analysis using an antibody specific for phosphocholine (TEPC-15; Sigma) to detect modified Rab1 and an antibody specific for Rab1A (Santa Cruz Biotechnologies) to detect total Rab1. Data shown were validated in three independent experiments.

Detection of host proteins modified by AnkX and CBU_2048 in vivo. HEK293 cells in 24-well dishes were transfected with plasmids encoding GFP–CBU_2078, GFP–AnkX or GFP–AnkX(H229A) and cultured for 18 h. Cells were lysed in buffer containing 20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM $MgCl_2$, 1% Triton X-100, 1 mM PMSF and protease inhibitor cocktail. The lysates were centrifuged at 17,900*g* and 50 μg of the supernatant was separated by SDS–PAGE for immunoblot analysis using the anti-phosphocholine-specific antibody TEPC-15 (Sigma).

GEF binding assays. To measure connecdenn binding to modified Rab35, a plasmid encoding 3×-Flag–Rab35(S22N) was co-transfected into cells together with a plasmid encoding GFP–CBU_2078, GFP–AnkX or GFP–AnkX(H229A). The GDP-locked allele of Rab35 was used because it demonstrates enhanced connecdenn binding. Cellular lysates were prepared as described earlier, and post-nuclear supernatant was incubated with 4 μg of purified connecdenn (GenBank accession number NP_659414). 3×-Flag–Rab35(S22N) was precipitated using anti-Flag agarose (Sigma) and eluted from the beads using the 3×-Flag peptide (Sigma). The eluted proteins were resolved by SDS–PAGE gel and proteins were identified after the gel was stained with Coomassie brilliant blue dye. The locations of 3×-Flag–Rab35(S22N) and connecdenn were determined by running purified connecdenn and 3×-Flag–Rab35(S22N) in adjacent wells. To measure DrrA binding to modified Rab1, HEK293 cells that stably produce 3×-Flag–Rab1 were transfected with GFP–CBU_2078, GFP–AnkX or GFP–AnkX(H229A). 500 μg of post-nuclear supernatant was incubated with 0.5 μg of purified His-tagged DrrA(340–533). 3×-Flag–Rab1 was precipitated using anti-Flag agarose beads (Sigma). Immunoblot analysis was used to compare the amount of DrrA (anti-His), the amount of total 3×-Flag–Rab1 (anti-Flag, Sigma) and the amount of phosphocholinated Rab1 (TEPC-15; Sigma) after 2% of the total amount of 3×-Flag–Rab1 precipitated from each reaction was fractionated by SDS–PAGE. Data shown were validated in three independent experiments.

Mass spectrometry. 3×-Flag–Rab proteins were immunoprecipitated from cells using anti-Flag agarose beads, and separated by SDS–PAGE. The band corresponding to the 3×-Flag–Rab protein was excised from the gel and treated with DTT to reduce disulphide bonds and then alkylated with iodoacetamide (IAM). Trypsin digestion was allowed to occur overnight. Resulting peptides were extracted from gel matrix and then resuspended in aqueous buffer before final LC-MS/MS analysis. Nanoflow reverse-phase LC separation was carried out on a Proxeon EASY-nLC system (Thermo Scientific). The capillary column (75 μm × 150 mm, PICOFRIT, New Objective) was packed in-house. A methanol slurry containing 5 μm, 100 Å Magic C18AQ silica-based particles (Microm BioResources) was forced to run through an empty capillary (with a frit in the end) using a pressurized device. The LC mobile phase was comprised of solvent A (97% H_2O, 3% acetonitrile (ACN) and 0.1% formic acid (FA)) and solvent B (100% ACN and 0.1% FA). The nano-LC separation was performed with the following gradient: B was increased from 7% to 35% in 40 min and then raised to 90% in 3 min and kept there for 10 min before going back to 100% A for column equilibration. At the moment when peptides were eluted from the capillary column, they were electrosprayed directly onto a linear ion trap mass spectrometer (LTQ Velos, ThermoElectron) for MS and MS/MS analysis. A data-dependent mode was enabled for peptide fragmentation with one full MS scan followed by collision induced dissociation (CID) of the ten most intense peptide ions. Dynamic exclusion was enabled to preclude repeated analyses of the same precursor ion. MS/MS scans were processed and searched using MASCOT (Matrix Science). The resulting peptide and protein assignments were filtered to keep only those identifications with scores above extensive homology. High-resolution MS/MS analysis was performed and the data was acquired by a high-resolution mass spectrometer (Orbitrap) by the Keck Proteomic Facility at Yale University. All LC-MS/MS data was validated by at least two independent experiments.

24. Kagan, J. C., Stein, M. P., Pypaert, M. & Roy, C. R. *Legionella* subvert the functions of Rab1 and Sec22b to create a replicative organelle. *J. Exp. Med.* **199,** 1201–1211 (2004).
25. Zuckman, D. M., Hung, J. B. & Roy, C. R. Pore-forming activity is not sufficient for *Legionella pneumophila* phagosome trafficking and intracellular growth. *Mol. Microbiol.* **32,** 990–1001 (1999).

26. Berger, K. H., Merriam, J. J. & Isberg, R. R. Altered intracellular targeting properties associated with mutations in the *Legionella pneumophila dotA* gene. *Mol. Microbiol.* **14,** 809–822 (1994).

27. Arasaki, K. & Roy, C. R. *Legionella pneumophila* promotes functional interactions between plasma membrane syntaxins and Sec22b. *Traffic* **11,** 587–600 (2010).

α–Synuclein occurs physiologically as a helically folded tetramer that resists aggregation

Tim Bartels[1], Joanna G. Choi[1] & Dennis J. Selkoe[1]

Parkinson's disease is the second most common neurodegenerative disorder[1,2]. Growing evidence indicates a causative role of misfolded forms of the protein α-synuclein in the pathogenesis of Parkinson's disease[3,4]. Intraneuronal aggregates of α-synuclein occur in Lewy bodies and Lewy neurites[5], the cytopathological hallmarks of Parkinson's disease and related disorders called synucleinopathies[4]. α-Synuclein has long been defined as a 'natively unfolded' monomer of about 14 kDa (ref. 6) that is believed to acquire α-helical secondary structure only upon binding to lipid vesicles[7]. This concept derives from the widespread use of recombinant bacterial expression protocols for *in vitro* studies, and of overexpression, sample heating and/or denaturing gels for cell culture and tissue studies. In contrast, we report that endogenous α-synuclein isolated and analysed under non-denaturing conditions from neuronal and non-neuronal cell lines, brain tissue and living human cells occurs in large part as a folded tetramer of about 58 kDa. Several methods, including analytical ultracentrifugation, scanning transmission electron microscopy and *in vitro* cell crosslinking confirmed the occurrence of the tetramer. Native, cell-derived α-synuclein showed α-helical structure without lipid addition and had much greater lipid-binding capacity than the recombinant α-synuclein studied heretofore. Whereas recombinantly expressed monomers readily aggregated into amyloid-like fibrils *in vitro*, native human tetramers underwent little or no amyloid-like aggregation. On the basis of these findings, we propose that destabilization of the helically folded tetramer precedes α-synuclein misfolding and aggregation in Parkinson's disease and other human synucleinopathies, and that small molecules that stabilize the physiological tetramer could reduce α-synuclein pathogenicity.

To identify the native state of α-synuclein in cells while avoiding the potential breakdown of physiological assemblies by detergents, we initially used native gel electrophoresis. α-Synuclein is expressed endogenously in many cell types, so we chose to analyse the dopaminergic human neuroblastoma line, M17D (ref. 8) and the commonly used cell lines HEK293, HeLa and COS-7. Each of these cell lines predominantly contained a non-denatured α-synuclein-immunoreactive species migrating in blue native–polyacrylamide gel electrophoresis (BN–PAGE) at ~45–50 kDa (Fig. 1a, lanes 1–4). Because these initial results suggested an apparently stable oligomeric form under native conditions, we next probed the endogenous state of α-synuclein in normal brain. The frontal cortex of wild-type mice also revealed a ~45–50 kDa form of endogenous α-synuclein as the main species in the buffer-soluble fraction (Fig. 1a, lane 6).

To assess the state of endogenous α-synuclein in living human cells, we examined freshly collected red blood cells (RBCs), which were recently found to have high α-synuclein expression[9]. Human RBCs contained a ~45–50 kDa α-synuclein immunoreactive band on BN–PAGE (Fig. 1a, lane 5). As a second non-denaturing gel system that precludes effects of the Coomassie dye used in BN–PAGE, we performed clear native–PAGE (CN–PAGE)[10]. The main α-synuclein species in all samples migrated at ~55–60 kDa, suggesting a tetramer (theoretical mass of monomer = 14,460 Da) (Fig. 1b, lanes 1–6). The

better resolution of CN–PAGE without Coomassie dye also revealed small amounts of apparent monomers running below the 14 kDa molecular weight marker (Fig. 1b, lanes 1–4, 6) and distinguished the small differences in amino acid length of the human and mouse α-synuclein monomers and putative tetramers (Fig. 1b, lane 6). The endogenous ~55–60 kDa species was detected by monoclonal α-synuclein antibodies Syn1, Syn211 and LB509 and polyclonal antibody C20 in both native gel systems.

Because the migration of a protein on BN– or CN–PAGE does not depend solely on its mass but also on its conformation and charge, we used *in vitro* crosslinking to preserve the assembled state of the putative α-synuclein oligomer, followed by denaturing SDS–PAGE. We observed SDS-stable α-synuclein bands migrating at the expected positions of a tetramer (~55 kDa) and non-crosslinked monomer in all
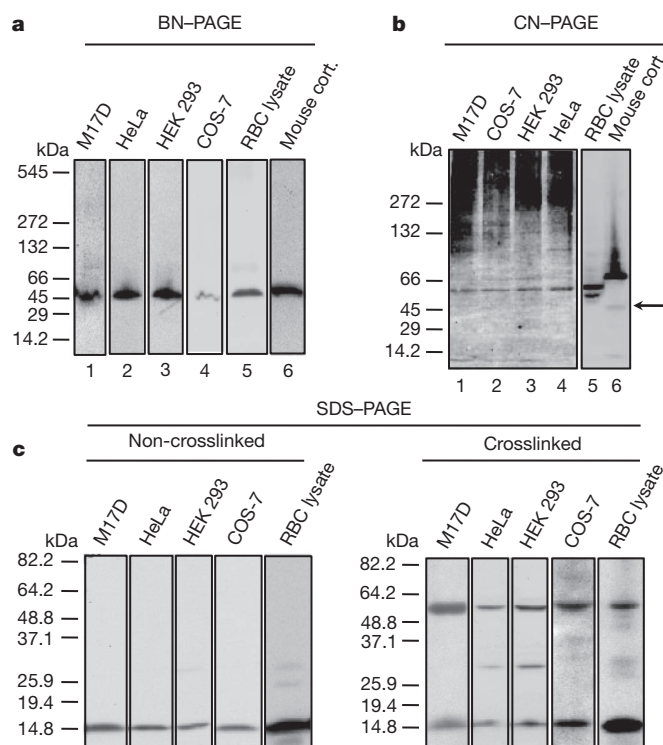
Figure 1 | Western blot analysis of lysates of M17D, HeLa, HEK293 and COS-7 cells, mouse cortex and human RBCs probed for endogenous α-synuclein. a, BN–PAGE. b, CN–PAGE. The band just below the main ~55–60 kDa RBC species (lane 6) may represent an alternatively spliced form of α-synuclein. Arrow marks a possible dimeric species. c, Left, SDS–PAGE/ western blot (antibody C20) analysis of cell lysates without crosslinking. Right, proteins were crosslinked in intact living cells with membrane permeable disuccinimidyl suberate (DSS) (M17D, HeLa, HEK 293, COS-7) or in RBC lysate with water soluble bis(sulphosuccinimidyl) suberate (BS³) and then run on SDS–PAGE.

[1]Center for Neurologic Diseases, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA.

cells, plus some putative dimer in the HeLa, HEK and red blood cells (right panel of Fig. 1c). This *in vivo* crosslinking supports the existence of native tetramers in cells. We performed two-dimensional gel analysis after the *in vivo* crosslinking; that is, isoelectric focusing (IEF) to separate proteins by charge in a pH gradient followed by denaturing SDS–PAGE. The higher-migrating α-synuclein species in the cross-linked RBC lysates had the same pK_a as monomers, within the limits of IEF resolution (Supplementary Fig. 1), consistent with their being homo-oligomers.

Next, we developed a non-denaturing method to purify native α-synuclein from soluble RBC lysates (see Methods and *Nature Protocol Exchange* (http://www.nature.com/protocolexchange/protocols/2136)). This allowed us to estimate the mass of native α-synuclein based on distinct measurement principles that are not affected by protein conformation, unlike gel electrophoresis. Scanning transmission electron microscopy (STEM) is useful for measuring the masses of purified, non-covalently bonded complexes that may not resist ionization during mass spectrometry[11,12]. STEM images of α-synuclein purified under non-denaturing conditions from human RBCs (Supplementary Fig. 2) yielded a homogenous distribution of roughly spherical particles measuring ~3.0–3.5 nm in diameter (Fig. 2a). Unbiased automatic sampling of 1,000 particles gave a size distribution pattern with a peak at ~55 kDa (Fig. 2b). Importantly, we next applied sedimentation equilibrium analytical ultracentrifugation (SE-AUC), a technique commonly used to establish the oligomeric state of native proteins independent of their conformation. SE-AUC analysis of purified, native RBC α-synuclein performed at three different concentrations and at different rotor speeds yielded an average molecular weight of 57.8 kDa (4.78 Svedbergs), strongly supporting a tetrameric assembly state (Fig. 2c).

Numerous studies have reported conformational changes in α-synuclein, with a focus on the natively unfolded recombinant monomer undergoing a random coil to α-helix transition upon *in vitro* interaction with small lipid vesicles[7]. This change is believed to be relevant to the poorly defined physiological function of α-synuclein in cells and could potentially decrease the likelihood of its aggregation into β-sheet-rich neurotoxic assemblies[13]. Unexpectedly, we found

that circular dichroism spectra of the human RBC tetramer purified under non-denaturing conditions showed two minima of mean residue ellipticity at 222 and 208 nm (Fig. 3a), characteristic of an α-helically folded protein[14]. This result is inconsistent with the common assumption that α-synuclein is natively unfolded. The addition of negatively charged, small unilamellar lipid vesicles (SUVs) did not induce a significant conformational change in the native tetramer by circular dichroism (Fig. 3a), but a random coil to α-helical conversion did occur (as reported) with recombinant monomer that had been expressed in bacteria (Fig. 3b). Incubation of the purified RBC α-synuclein tetramer with Lipidex 1000, a reagent used to strip proteins of bound lipids and fatty acids[15], did not change the conformation of the α-helical α-synuclein tetramer (Supplementary Fig. 3), suggesting that significant lipid association is not required to maintain the folded structure of cellular α-synuclein. To support this possibility, we conducted a quantitative elemental phosphate analysis[16] on the purified native α-synuclein to search for phospholipid. We obtained an average value of 0.25 mol phosphate per mol α-synuclein, making a significant presence of phospholipids on the α-helical α-synuclein purified from normal cells unlikely. Because post-translational modifications also could have an impact on the conformational differences between the native human RBC tetramer and the bacterially expressed, recombinant human monomer, we performed mass spectrometry. The recombinant protein showed a mass peak at 14,462 kDa, very close to the theoretical predicted mass of 14,460 kDa, whereas the purified erythrocyte α-synuclein showed a peak at 14,505 kDa, indicative of only an N-α-acetylation commonly present on human proteins (theoretical predicted mass = 14,502 kDa) (Supplementary Fig. 4).

To validate the above results obtained on RBC α-synuclein using a different human cell type and a different non-denaturing purification method, we isolated α-synuclein from a M17D human neuroblastoma cell line stably overexpressing wild-type human α-synuclein (3D5 cells[17]). α-Synuclein from untransfected M17D cell lysates migrated above bacterially expressed α-synuclein of confirmed random coil structure on CN–PAGE (Supplementary Fig. 5A). This higher electrophoretic migration was also true of native (α-helical) but not
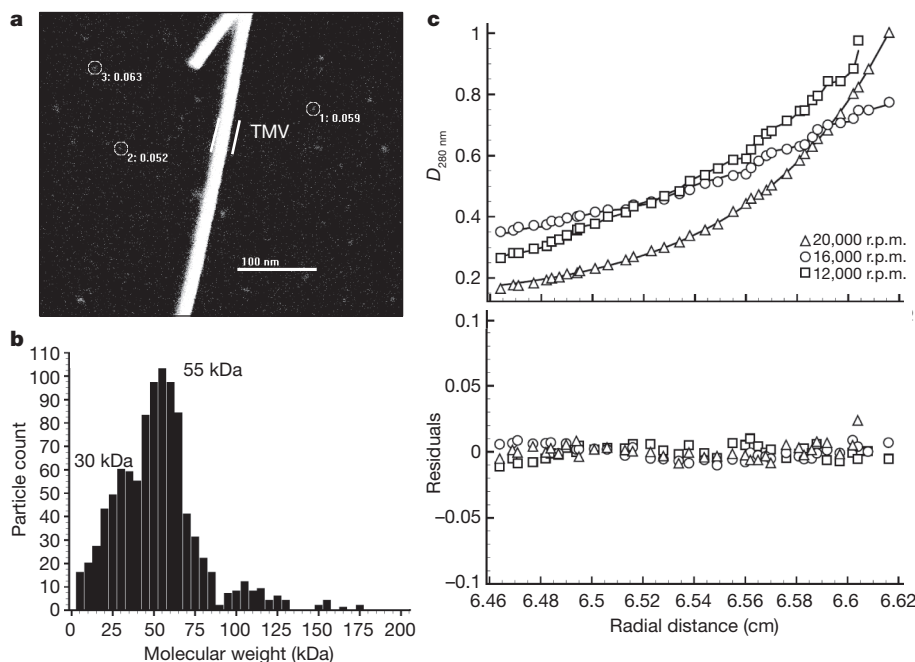
Figure 2 | Sizing analyses of α-synuclein from human RBCs.
a, Representative large-angle dark-field STEM image of purified α-synuclein from human RBCs. A few representative particles are circled. As an internal size standard, tobacco mosaic virus (TMV) helical rod was included during electron microscopy specimen preparation. **b**, Mass histogram (bin size = 5 kDa) of 1,000 automatically selected α-synuclein particles. **c**, SE-AUC of purified,

native RBC α-synuclein. Top panel shows the individual experimental analyses fitting an ideal single-species model to the equilibrium data obtained at 11,612, 20,644 and 32,256g for 1.1 mg ml^{-1} α-synuclein solution. The fitting yielded a molecular weight of 57,753 Da (standard deviation (s.d.) ± 655.199) with a root mean squared deviation of 0.004533. *D*, attenance. Bottom panel shows an overlay of the residuals of data and theoretical fit for the three different speeds.

denatured (random coil) purified RBC α-synuclein (Supplementary Fig. 5B). After α-synuclein was purified from the stably transfected 3D5 cell line or from RBCs, the two differentially purified and α-helically folded (by circular dichroism) cellular proteins co-migrated at ~55–60 kDa on CN–PAGE, as expected (Supplementary Fig. 6). Unbiased, automated STEM measurements of 3,000 particles revealed that the 3D5 neuroblastoma cells contained α-synuclein tetramers of closely similar estimated molecular weight (peak mass ~55 kDa) to those of the RBC α-synuclein (Supplementary Fig. 7; compare to Fig. 2b). Circular dichroism spectroscopy revealed the purified 3D5 cell α-synuclein to have two minima of mean residue ellipticity at 222 and 208 nm (Supplementary Fig. 8). To further exclude artefacts arising during purification of cellular α-synuclein such as adventitious association of biomolecules (for example, cellular lipids not removed by Lipidex 1000) that artificially fold the protein, we repeated our experiments with the 3D5 parental line M17D, which has only low levels of endogenous α-synuclein. We added ('spiked') bacterially expressed recombinant human monomer onto the M17D cells before performing cell lysis and the full purification, and then assayed its structural properties. This exposure to cell lysates and the purification procedure led to no induction of helical folding in the recombinant human α-synuclein (Supplementary Fig. 9), whereas simultaneously purified 3D5 cell human α-synuclein did show this conformation, supporting our conclusion that α-helically folded α-synuclein does not arise owing to artificial manipulation of the protein.

Membrane association has been viewed as a principal functional property of α-synuclein *in vitro*[7] and in living cells[18]. We searched for differential binding of recombinant monomeric human α-synuclein versus RBC tetrameric human α-synuclein to a lipid membrane using surface plasmon resonance (SPR). Because recombinant α-synuclein is reported to have preferential affinity for negatively charged lipids, especially phosphatidyl serine[7], we chose a mixed phosphatidyl choline and phosphatidyl serine (PC/PS) membrane as a model membrane. Exposure of a PC/PS membrane to cell-derived, purified native α-synuclein in a Biacore instrument produced a markedly increased resonance angle shift compared to conventional recombinant monomers at identical concentrations in solution (Fig. 3c), indicating dramatically

increased lipid binding. Fitting a dilution series of α-synuclein tetramer injections to a two-state binding model (Supplementary Fig. 10) gave an apparent dissociation constant of $K_{app} = 56 \pm 61$ nM, which is about two orders of magnitude lower than values obtained for recombinant monomer in analogous SPR studies[19]. We next tested the amyloid aggregation propensity of the distinct species in a Thioflavin T (ThT) fluorescence assay. Monomeric and tetrameric α-synuclein showed very different characteristics, with samples of purified cellular α-synuclein incubated under identical conditions showing no evidence of fibril formation in a time (10 days) more than sufficient to form mature, ThT-bound fibrils from equivalent amounts of unfolded recombinant α-synuclein (Fig. 3d). Analysis of protein concentration in the solution after the 10-day incubation showed that the RBC α-synuclein was still present and soluble, making non-amyloid (that is, ThT-negative) aggregation of the tetramers unlikely. Interestingly, melting curves of purified tetrameric α-synuclein showed that heat denaturation (at 95 °C) seemed irreversible under our conditions (Supplementary Fig. 11).

Our experiments provide several independent lines of evidence that endogenous cellular α-synuclein exists in large part as an α-helically folded, ~58 kDa tetramer under native conditions. This finding is in contrast to many biophysical and biochemical studies describing α-synuclein as a natively unfolded ~14 kDa monomer. In an early study of bacterially expressed recombinant protein purified under non-denaturing conditions or with heat treatment, no conformational differences were observed, and it was concluded that α-synuclein is a natively unfolded monomer[6]. This suggests problems in generating properly folded protein in Escherichia coli, although a modified bacterial expression protocol avoiding heating and denaturants has recently been found to yield a helical α-synuclein tetramer closely resembling the species found by us in native human samples (W. Wang *et al.*, personal communication). The reasons for the conformational differences observed in these two bacterial studies are unknown. Using gel filtration on unfolded recombinant α-synuclein also showed an apparent molecular weight of ~60 kDa in some earlier studies; the data were interpreted as a decrease in mobility of the extended state of an unfolded protein in the tested matrices[6]. This suggests the possibility
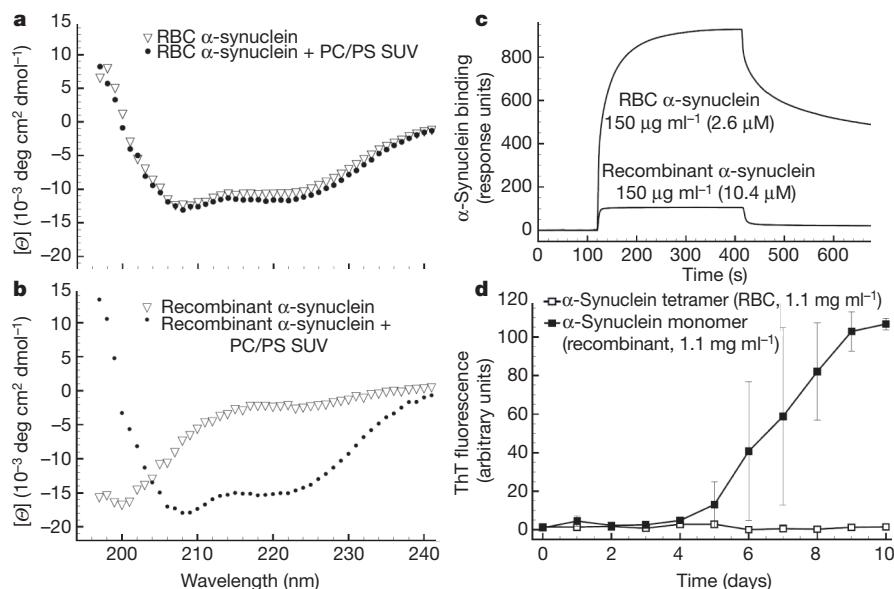
Figure 3 | Comparative analyses of native (cell-derived) and bacterial α-synuclein. a, Circular dichroism spectra of native tetrameric α-synuclein (isolated under non-denaturing conditions from human RBCs) before versus after addition of POPC/POPS SUVs (PC/PS 4:1; protein:lipid 1:500). **b,** Circular dichroism spectra of recombinant α-synuclein monomer purified from *E. coli*, alone and with addition of PC/PS SUVs (protein/lipid 1:500). **c,** SPR sensorgram of equal protein concentrations of α-synuclein

recombinant monomer versus endogenous tetramer injected on a L1 chip covered with a PC/PS membrane. **d,** Amyloid-type aggregation kinetics of recombinant α-synuclein monomer versus native RBC tetramer monitored by ThT fluorescence; average values from 3 independent experiments (error bars show s.d.; some s.d. values for RBC-derived α-synuclein are smaller than the symbol size).

of a similar hydrodynamic radius for the unfolded monomer and the more compact, helically folded tetramer, making gel filtration an unreliable indicator (and therefore not used here). Our evidence for a tetrameric molecular mass of endogenous α-synuclein was particularly supported by the analytical ultracentrifugation and the unbiased STEM analysis, both of which sizing methods are not based on conformation. The STEM sizing was performed on intrinsic α-synuclein isolated from two cell types and using two distinct non-denaturing procedures.

Our apparent disagreement with most published findings on the monomeric state of α-synuclein in cells and brain tissue, usually as judged by SDS–PAGE and western blotting, can be explained by the common use of denaturing detergents. Our tetramer aggregation data (Fig. 3d) are consistent with a recent report describing non-neurotoxic, aggregation-resistant α-synuclein oligomers *in vivo*[20]. Moreover, an oligomeric species of α-synuclein (size undefined) was observed by *in vivo* fluorescence lifetime imaging in an intact cell culture model[21]. Given the close match between our observed molecular weights using SE-AUC (Fig. 2c) and STEM (Fig. 2b and Supplementary Fig. 7) and the theoretical weight of a tetramer, the detection of a tetrameric band on denaturing gels after *in vivo* crosslinking (Fig. 1c), and the IEF evidence post-crosslinking that the endogenous tetramer and dimer bands have pK_a values similar to that of a monomer (Supplementary Fig. 1), we conclude that the predominant physiological species of α-synuclein in cells and brain is a helically folded tetramer, although minor and variable amounts of monomers, dimers and trimers were detected in some cell types. The closely similar properties of α-synuclein observed until now in neural cells and fresh human RBCs recommends the latter as an abundant, available source for future studies of physiological α-synuclein.

The higher lipid-binding capacity of native α-synuclein leads us to speculate that the monomer represents a not fully functional and less abundant species in normal cells. Given the much lower propensity of the native tetramer to aggregate into fibrils (Fig. 3d), it is likely that tetramers undergo destabilization before α-synuclein aggregates into abnormal oligomeric and fibrillar assemblies that can confer cytotoxicity in Parkinson's disease and other α-synucleinopathies. Hypothetically, such a mechanism could be analogous in part to transthyretin amyloidosis, in which a native metastable tetramer circulates in human plasma but can become destabilized (for example, by pathogenic missense mutations) to allow monomers to aggregate aberrantly in tissue[22]. Our identification of helically folded α-synuclein tetramers encourages the design of compounds that, like those for transthyretin[23], could kinetically stabilize native tetramers and prevent pathogenic α-synuclein aggregation as a novel treatment approach for Parkinson's disease, dementia with Lewy bodies and other synucleinopathies[24].

METHODS SUMMARY

Native gel electrophoresis was conducted as described[10]. For crosslinking, 1–5 mM disuccinimidyl suberate was added to living cells. RBC lysates were treated analogously but using 1 mM bis(sulphosuccinimidyl) suberate. To purify α-synuclein from fresh or packed frozen RBCs, an initial 25% $(NH_4)_2SO_4$ cut followed by a 50% $(NH_4)_2SO_4$ precipitation substantially enriched α-synuclein. The resolubilized 50% pellet was injected onto a hydrophobic interaction column (HiTrap Phenyl HP; GE Healthcare) and eluted in 1 M to 0 M $(NH_4)_2SO_4$, pH 7. Alternatively, α-synuclein-overexpressing 3D5 neuroblastoma cell lysate after $(NH_4)_2SO_4$ was injected onto a 5-ml HiTrap Q HP column. A 25–500 mM NaCl (pH 8.0) gradient eluted α-synuclein at ~300 mM NaCl. α-Synuclein from both cell sources underwent a final purification step on a Superdex 75 SEC column. STEM analysis was conducted at the Brookhaven National Laboratory STEM user facility. Sedimentation equilibrium data were acquired on a Beckman XL-I analytical ultracentrifuge at speeds of 11,612, 20,644 and 32,256g (AN-60 Ti rotor) and protein concentrations of 0.6, 1.1 and 1.6 mg ml^{-1}. Circular dichroism spectroscopy for lipid-induced α-synuclein folding was conducted in the presence of 4 mM PC/PS (4:1) SUVs. SPR spectroscopy was conducted as described[19]. To quantify amyloid fibril growth, aliquots (10 μl) of purified α-synuclein were added to a 10 μM ThT solution in 10 mM glycine buffer, pH 9. ThT fluorescence was measured by exciting at 444 nm and scanning the emission wavelengths from 460–550 nm.

1. Obeso, J. A. *et al.* Missing pieces in the Parkinson's disease puzzle. *Nature Med.* **16**, 653–661 (2010).
2. Gupta, A., Dawson, V. L. & Dawson, T. M. What causes cell death in Parkinson's disease? *Ann. Neurol.* **64**, S3–S15 (2008).
3. Winklhofer, K. F., Tatzelt, J. & Haass, C. The two faces of protein misfolding: gain- and loss-of-function in neurodegenerative diseases. *EMBO J.* **27**, 336–349 (2008).
4. Tong, J. *et al.* Brain α-synuclein accumulation in multiple system atrophy, Parkinson's disease and progressive supranuclear palsy: a comparative investigation. *Brain* **133**, 172–188 (2010).
5. Spillantini, M. G. *et al.* α-Synuclein in Lewy bodies. *Nature* **388**, 839–840 (1997).
6. Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A. & Lansbury, P. T. J. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* **35**, 13709–13715 (1996).
7. Davidson, W. S., Jonas, A., Clayton, D. F. & George, J. M. Stabilization of α-synuclein secondary structure upon binding to synthetic membranes. *J. Biol. Chem.* **273**, 9443–9449 (1998).
8. DeTure, M. *et al.* Missense tau mutations identified in FTDP-17 have a small effect on tau-microtubule interactions. *Brain Res.* **853**, 5–14 (2000).
9. Scherzer, C. R. *et al.* GATA transcription factors directly regulate the Parkinson's disease-linked gene α-synuclein. *Proc. Natl Acad. Sci. USA* **105**, 10907–10912 (2008).
10. Wittig, I. & Schagger, H. Advantages and limitations of clear-native PAGE. *Proteomics* **5**, 4338–4346 (2005).
11. Osenkowski, P. *et al.* Cryoelectron microscopy structure of purified •-secretase at 12 Å resolution. *J. Mol. Biol.* **385**, 642–652 (2009).
12. Wall, J. S., Simon, M. N., Lin, B. Y. & Vinogradov, S. N. Mass mapping of large globin complexes by scanning transmission electron microscopy. *Methods Enzymol.* **436**, 487–501 (2008).
13. Beyer, K. Mechanistic aspects of Parkinson's disease: α-synuclein and the biomembrane. *Cell Biochem. Biophys.* **47**, 285–299 (2007).
14. Chen, Y., Yang, J. T. & Martinez, H. M. Determination of the secondary structures of proteins by circular dichroism and optical rotatory dispersion. *Biochemistry* **11**, 4120–4131 (1972).
15. Sharon, R. *et al.* α-Synuclein occurs in lipid-rich high molecular weight complexes, binds fatty acids, and shows homology to the fatty acid-binding proteins. *Proc. Natl Acad. Sci. USA* **98**, 9110–9115 (2001).
16. Chen, P. S., Toribara, T. Y. & Warner, H. Microdetermination of phosphorus. *Anal. Chem.* **28**, 1756–1758 (1956).
17. Ko, L. W., Ko, H. H., Lin, W. L., Kulathingal, J. G. & Yen, S. H. Aggregates assembled from overexpression of wild-type α-synuclein are not toxic to human neuronal cells. *J. Neuropathol. Exp. Neurol.* **67**, 1084–1096 (2008).
18. McLean, P. J., Kawamata, H., Ribich, S. & Hyman, B. T. Membrane association and protein conformation of α-synuclein in intact neurons. Effect of Parkinson's disease-linked mutations. *J. Biol. Chem.* **275**, 8812–8816 (2000).
19. Smith, D. P. *et al.* Formation of a high affinity lipid-binding intermediate during the early aggregation phase of α-synuclein. *Biochemistry* **47**, 1425–1434 (2008).
20. Tsika, E. *et al.* Distinct region-specific α-synuclein oligomers in A53T transgenic mice: implications for neurodegeneration. *J. Neurosci.* **30**, 3409–3418 (2010).
21. Klucken, J., Outeiro, T. F., Nguyen, P., McLean, P. J. & Hyman, B. T. Detection of novel intracellular α-synuclein oligomeric species by fluorescence lifetime imaging. *FASEB J.* **20**, 2050–2057 (2006).
22. Quintas, A., Saraiva, M. J. M. & Brito, R. M. M. The tetrameric protein transthyretin dissociates to a non-native monomer in solution. *J. Biol. Chem.* **274**, 32943–32949 (1999).
23. Connelly, S., Choi, S., Johnson, S. M., Kelly, J. W. & Wilson, I. A. Structure-based design of kinetic stabilizers that ameliorate the transthyretin amyloidoses. *Curr. Opin. Struct. Biol.* **20**, 54–62 (2010).
24. Lansbury, P. T. & Lashuel, H. A. A century-old debate on protein aggregation and neurodegeneration enters the clinic. *Nature* **443**, 774–779 (2006).

METHODS

Materials. Recombinant human α-synuclein was bought from Anaspec. Recombinant human transthyretin was provided by I. Rappley and J. Kelly (Scripps Research Institute). HEK, COS-7 and HeLa cells were cultured in DMEM with 10% fetal bovine serum, penicillin (100 U ml^{-1}), streptomycin (100 μg ml^{-1}) and L-glutamine (2 mM). For M17D and 3D5 human neuroblastoma cells, standard DMEM was supplemented with 400 μg ml^{-1} G418 and 1 μg ml^{-1} puromycin. Frontal cortex was obtained from wild-type mice aged 4–9 months. α-Synuclein immunoblotting used antibodies C20 (1:1,000, Santa Cruz), LB509 (1:400, Santa Cruz), Syn211 (1:200, Santa Cruz) and Syn1 (1:2,000, BD).

Lipid preparation. PC/PS SUVs (30 nm) of 80% 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and 20% 1-palmitoyl-2-oleoyl-sn-glycero-3-[phospho-L-serine] (POPS) (Avanti Polar Lipids) were prepared in 10 mM sodium phosphate, pH 7.4, by sonication.

Crosslinking. Cells were detached and incubated at room temperature (22 °C) for 30 min in DSS crosslinker (1–5 mM), then quenched with 1 M Tris buffer, pH 7.4, for 15 min at room temperature. Human RBC lysates were treated analogously with 1 mM BS^3 (Pierce) to covalently crosslink lysine residues.

BN–PAGE and CN–PAGE. For BN–PAGE, samples were run on 4–16% Bis-Tris BN–PAGE gels (Invitrogen) at 200 V and room temperature. The cathode buffer was 50 mM tricine, 15 mM Bis-Tris, 0.02% Brilliant Blue G (Serva), pH 7.0; the anode buffer was 50 mM Bis-Tris pH 7.0. CN–PAGE was conducted identically to BN–PAGE, but Coomassie blue was omitted from the sample and the cathode buffer. Electroblotting of protein on PVDF membranes (0.45 μm pore size) was conducted at 400 mA for 2 h. For molecular weight estimation three different molecular weight marker were loaded on each gel (Sigma Non-denaturing, 108K6408, Invitrogen Native Mark, LC0725 GE Healthcare HMW Native Marker Kit, 17-0445-01).

IEF two-dimensional PAGE. We used the IPGphor IEF system (GE Healthcare). Lysates were heated at 65 °C overnight and brought to 200 μl with sample rehydration buffer (7 M urea, 2 M thiourea, 2% Chapso, 0.5% IPG buffer (GE Healthcare, bromophenol blue) and applied on an 11 cm 1D Ready-strip (Bio-Rad) with a pH gradient of 4–7. Sample was rehydrated for 16 h followed by IEF at 500 V for 30 min, then 1,000 V for another 30 min, and then 8,000 V for 3.5 h. The 1D strip was then applied to a precast NuPAGE ZOOM 4–12% Bis-Tris gel (Invitrogen) and run at 200 V.

Purification of α-synuclein from human RBCs. Freshly collected and washed RBCs were resuspended in a threefold volume of ACK lysing buffer (Lonza). (NH$_4$)$_2$SO$_4$ to a final concentration of 25% was added and incubated at 4 °C for 30 min. The lysate was centrifuged (20,000g, 20 min), and the supernatant brought up to 50% (NH$_4$)$_2$SO$_4$. The pellet was washed several time in 55% (NH$_4$)$_2$SO$_4$ to remove excess haemoglobin. The sample was centrifuged at 20,000g for 20 min and the pellet resolubilized in a 50-fold volume of 50 mM phosphate buffer, pH 7.0, 1 M (NH$_4$)$_2$SO$_4$. Five millilitres of the resultant solution were injected onto a 5-ml HiTrap phenyl hydrophobic interaction column (GE Healthcare) equilibrated with 50 mM phosphate buffer, pH 7.0, 1 M (NH$_4$)$_2$SO$_4$. α-Synuclein was eluted with a 1 M to 0 M (NH$_4$)$_2$SO$_4$ gradient in 50 mM phosphate buffer, pH 7.0 (α-synuclein eluted at ~0.75 M (NH$_4$)$_2$SO$_4$). For anion exchange purification of RBC α-synuclein, we used the protocol used for neuroblastoma cells (see later), but the first run of RBC lysate sometimes showed low binding of α-synuclein and contamination by plasma transthyretin. In these cases we discarded the first eluate and used the flow-through for a second run, which showed significantly higher binding and subsequent purity. As a third alternative to HIC and AX, an XK 16/100 column packed with activated thiolpropyl Sepharose 6B gel media was used (Supplementary Fig. 5B) (binding buffer, PBS; flow rate, 0.2 ml min^{-1}). In this case, the flow-trough contained α-synuclein and was processed further. The column was regenerated by eluting bound protein with 5 column volumes of binding buffer with 25 mM dithiothreitol, and reactivated with 1.5 mM dipyridyl sulphide in 50 mM borate buffer pH 8.0. The final solution was concentrated (Amicon Ultra centrifugal filter units, MWCO 10,000, Millipore) and further purified via gel filtration.

Purification of α-synuclein from human neuroblastoma cells. 3D5 cells (α-synuclein stables) and their parental M17D cells were scraped from the plates, washed in PBS and lysed by sonication. A (NH$_4$)$_2$SO$_4$ precipitation was conducted as described earlier, the 50% (NH$_4$)$_2$SO$_4$ pellet was taken up in 20 mM tris buffer, pH 8.0, 25 mM NaCl. The sample was injected onto a 5-ml HiTrap Q HP anion exchange column, equilibrated with 20 mM Tris buffer, pH 8.0, 25 mM NaCl.

α-Synuclein was eluted from the column with a 25–500 mM NaCl gradient in 20 mM Tris buffer, pH 8.0. α-Synuclein was eluted at ~300 mM NaCl. The column was regenerated with 1 M NaCl in 20 mM tris buffer, pH 8.0. The final solution was concentrated (Amicon Ultra centrifugal filter units, MWCO 10,000, Millipore) and further purified via gel filtration. For the addition ('spiking') of exogenous recombinant α-synuclein monomers, bacterially expressed α-synuclein was added to the scraped M17D cell pellet and the purification scheme conducted as just described.

Gel filtration. Aliquots (250 μl) were injected onto either a Superdex 75 (10/300 GL), Superdex 200 (10/300 GL) or a Superose 12 (10/300 GL) column (GE Healthcare) at 4 °C and eluted with 50 mM ammonium acetate, pH 8.5. For size estimation, a gel filtration standard (Bio-Rad, cat. no. 151-1901) was run on each column, and the calibration curve was obtained by semi-logarithmic plotting of molecular weight versus elution volume divided by void volume.

STEM. STEM was carried out at the Brookhaven National Laboratory STEM user facility with 100 μl of sample at a concentration of 300 μg ml^{-1} in 50 mM ammonium acetate, pH 7.4, and diluted to find the appropriate concentration for a homogenous particle distribution. Tobacco mosaic virus (TMV) rods were included during specimen preparation as an internal sizing standard.

Circular dichroism spectroscopy. Circular dichroism spectra were obtained using an Aviv Biomedical spectrometer (model 410) in the presence or absence of 4 mM PC/PS SUVs. The spectral contributions of buffer and SUVs were subtracted. Data are reported as mean residue ellipticities measured at 20 °C and a pathlength of 0.1 mm.

Lipidex 1000 treatment. 10% (w/v) Lipidex 1000 (Perkin Elmer) was washed with 50% methanol-ultra pure water and added to a 100 μM solution of purified α-synuclein from RBCs. The samples were stirred overnight at 37 °C, and α-synuclein was purified from that mixture via size exclusion chromatography.

SPR. All lipid binding experiments were performed at 20 °C on a BIACORE 3000 apparatus using the L1 sensor chip (Biacore AB). The running buffer was 10 mM sodium phosphate, pH 7.4. SUVs were applied to the sensor chip surface at a flow rate of 10 μl min^{-1} in the presence of 0.1 mM NaCl. Injections were done at a flow rate of 10 μl min^{-1} with 50 μl sample volume. Apparent K_d values were calculated from equilibrium data of several dilution series, collected at 300–320 s.

ThT binding. To detect amyloid fibril growth, a discontinuous assay was used. Aliquots (10 μl) were removed from each purified α-synuclein sample (lyophilized from 50 mM ammonium acetate, pH 7.4, and agitated at 37 °C at a concentration of 75 μM in 20 mM Bis-Tris propane, 100 mM LiCl, pH 7.4) and added to 2 ml of a 10 μM ThT solution in 10 mM glycine buffer, pH 9. Fluorescence was directly quantified on a Varian Eclipse fluorescence spectrophotometer at 20 °C by exciting at 444 nm and scanning the emission wavelengths from 460–550 nm with slit widths set at 5 nm (PMT at 750 V).

Quantitative phosphate analysis. Samples (2 × 15 μl and 2 × 30 μl of 1 mg ml^{-1} α-synuclein in 50 mM ammonium acetate, pH 7.4) were placed at the bottom of glass test tubes, 225 μl of 8.9 N H$_2$SO$_4$ (in deionized water) was added, and the mixture was heated for 25 min at 200–215 °C. Next, 75 μl H$_2$O$_2$ was added to all tubes at room temperature. After heating for 30 min at 200–215 °C, 1.95 ml deionized water and then 0.25 ml 2.5% ammonium molybdate(VI) tetrahydrate solution (in deionized water) were added at room temperature. After addition of 0.25 ml 10% ascorbic acid solution (in deionized water), the tubes were heated for 7 min at 100 °C, and samples were allowed to cool to room temperature. Absorbance at 820 nm was measured, and phosphate concentration calculated using a calibration curve obtained from 7 phosphate standard solutions ranging from 0–50 nmol phosphate (Sigma-Aldrich).

SE-AUC. AUC experiments were performed in a Beckman Optima XL-I analytical ultracentrifuge. Sedimentation equilibrium experiments were carried out at purified α-synuclein protein concentrations of 1.6, 1.1 and 0.6 mg ml^{-1} in 50 mM ammonium acetate, pH 8.5. The experiments were performed at 20 °C at 11,612, 20,644 and 32,256g (AN-60 Ti rotor), and data were collected at 278 nm. The software SEDPHAT (version 6.5) was used to calculate the M and s of the species present in equilibrium in the samples. For molecular weight analysis, we used the model 'Species Analysis' available in the SEDPHAT program with RI noise baseline correction. Analysis was performed for each protein concentration separately, and the molecular weight determined from the average obtained for the analyses of the 3 protein concentrations. The average errors and standard deviations were calculated using Monte-Carlo simulation, with 1,000 iterations and a confidence level of 0.68.

Solution structure of a minor and transiently formed state of a T4 lysozyme mutant

Guillaume Bouvignies[1,2,3]*, Pramodh Vallurupalli[1,2,3]*, D. Flemming Hansen[1,2,3], Bruno E. Correia[4,5], Oliver Lange[4], Alaji Bah[6], Robert M. Vernon[4,6], Frederick W. Dahlquist[7], David Baker[4] & Lewis E. Kay[1,2,3,6]

Proteins are inherently plastic molecules, whose function often critically depends on excursions between different molecular conformations (conformers)[1–3]. However, a rigorous understanding of the relation between a protein's structure, dynamics and function remains elusive. This is because many of the conformers on its energy landscape are only transiently formed and marginally populated (less than a few per cent of the total number of molecules), so that they cannot be individually characterized by most biophysical tools. Here we study a lysozyme mutant from phage T4 that binds hydrophobic molecules[4] and populates an excited state transiently (about 1 ms) to about 3% at 25 °C (ref. 5). We show that such binding occurs only via the ground state, and present the atomic-level model of the 'invisible', excited state obtained using a combined strategy of relaxation-dispersion NMR (ref. 6) and CS-Rosetta[7] model building that rationalizes this observation. The model was tested using structure-based design calculations identifying point mutants predicted to stabilize the excited state relative to the ground state. In this way a pair of mutations were introduced, inverting the relative populations of the ground and excited states and altering function. Our results suggest a mechanism for the evolution of a protein's function by changing the delicate balance between the states on its energy landscape. More generally, they show that our approach can generate and validate models of excited protein states.

A detailed characterization of the conformers along a protein's energy landscape is important for understanding the structure–function relationship and also because such an analysis provides insight into fundamental aspects of protein structure and dynamics. In this vein, numerous detailed studies of mutant lysozymes and lysozyme complexes from phage T4 have greatly increased our understanding of the inter-relation between structure, stability, folding and motion in proteins[8]. Among the approximately 700 mutant lysozymes and lysozyme complexes that have been characterized is a family where each member contains an engineered cavity in its hydrophobic core, generated by replacing larger amino acids with alanine (ref. 9). The point mutant causing the most pronounced stability change involved the replacement of a leucine at position 99 (referred to in what follows as L99A T4L), creating a cavity of $\sim$150 Å^3 in the carboxy terminus of the enzyme that is able to bind hydrophobic ligands[4]. Interestingly, X-ray studies showed that the L99A mutant undergoes the least rearrangement at the site of mutation, with the structure essentially unchanged[9].

Despite the fact that the wild-type and L99A T4L structures are virtually identical in the crystalline state, solution NMR studies of the L99A mutant indicated that many of the peaks were significantly broadened relative to the corresponding resonances in data sets recorded of the wild-type protein[5]. Spectral broadening is indicative of dynamics on the microsecond–millisecond timescale[6], and in this

case provides a clear indication that cavity creation introduces one or more dynamic processes that are not observed in the wild-type enzyme. Such dynamics can be studied by NMR transverse spin-relaxation experiments, in which the relaxation rates of probe nuclei are measured as a function of the strength of applied radio-frequency fields[6,10]. These experiments provide a powerful approach to quantify structural transitions in proteins because they are sensitive to microsecond–millisecond exchange processes in which a highly populated ground state (G) interconverts with conformers that can have much lower populations ($>$0.5%), referred to in what follows as excited states (E).

Initial [15]N Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion NMR experiments indicated that L99A T4L undergoes a dynamic process involving residues that are proximal to the cavity. The relaxation data were well fitted to a model in which a highly populated ground state (97%, 25 °C) interconverts with a second state that because of its low population (3%) and short lifetime ($\sim$1 ms) is 'invisible' in NMR spectra[5]. Using recently developed CPMG dispersion experiments[11], we have obtained nearly all of the backbone ^{1}H, ^{15}N and ^{13}C chemical shifts—as well as side-chain methyl ^{13}C chemical shifts—of the invisible excited state with a high level of accuracy (Supplementary Fig. 1, Supplementary Tables 1–5). Such chemical shifts are powerful constraints in structure calculations; when combined with computational protocols[7,12] they can be used to calculate accurate folds of small proteins, even in the absence of additional information, such as internuclear distances[13,14].

A comparison of the chemical shifts of the ground and excited states shows that conformational rearrangements occur in the vicinity of the cavity involving the C-terminal region of helix E and helices F, G, H and I (Fig. 1a, b). These regions do not become disordered in the excited state, as calculated squared order parameters reporting on the amplitudes of backbone motion from chemical shifts[15], S^2_{RCI}, change little between states (Fig. 1c). However, a decrease in helix propensity is noted for the C-terminal region of helix E, with a very significant concomitant increase in the helix content for the loop connecting helices F and G (Fig. 1d).

The ^{15}N, [1]H^N, [1]H$^\alpha$, [13]C$^\alpha$ and [13]C$'$ chemical shifts of the excited state were used to guide Rosetta 'loop' building and refinement[16] to generate structural models of the excited state (described in Supplementary Information, Supplementary Fig. 2). Only regions with significant chemical shift changes (residues 100–120, 132–146; Fig. 1a) were allowed to deviate from the X-ray structure of the L99A cavity mutant. The CS-Rosetta-based excited state conformers so produced are well converged, with pair-wise backbone root-mean-squared deviations (r.m.s.d.) for ten representative, low energy structures of 0.7 ± 0.2 Å over the region that was allowed to vary in the calculations (Supplementary Table 6). As a control, an identical protocol was used to generate the structure of the ground state in the vicinity of the cavity mutant, based on the same number of chemical shifts as for the excited

[1]Department of Molecular Genetics, The University of Toronto, Toronto, Ontario M5S 1A8, Canada. [2]Department of Biochemistry, The University of Toronto, Toronto, Ontario M5S 1A8, Canada. [3]Department of Chemistry, The University of Toronto, Toronto, Ontario M5S 1A8, Canada. [4]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. [5]Program in Computational Biology, Instituto Gulbenkian de Ciência, P-2780-156 Oeiras, Portugal. [6]Hospital for Sick Children, Program in Molecular Structure and Function, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [7]Department of Chemistry and Biochemistry, University of California Santa Barbara, Santa Barbara, California 93106, USA.
*These authors contributed equally to this work.
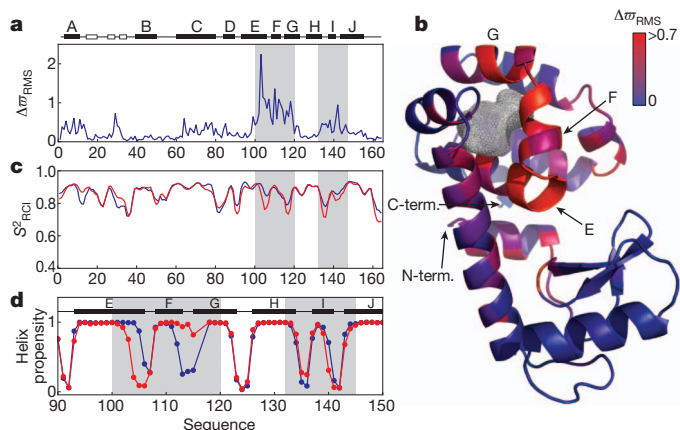
Figure 1 | L99A T4L exchanges between ground (visible) and excited (invisible) states, each with distinct conformations. **a**, Plot of

$$\Delta\varpi_{RMS} = \sqrt{\frac{1}{N}\sum_i \left(\frac{\Delta\varpi_i}{\Delta\varpi_{i,STD}}\right)^2}$$ as a function of residue, where $\Delta\varpi_i$ is the shift

difference in p.p.m. between states, $\Delta\varpi_{i,STD}$ is a nucleus specific value that corresponds to the range of shift values (1 s.d.) that are observed in a database of protein chemical shifts (http://www.bmrb.wisc.edu) for the nucleus in question ($^1H^N$, ^{15}N, $^{13}C^\alpha$, $^1H^\alpha$ and $^{13}C'$) and N is the number of nuclei ≤ 5 that are included in the average. Significant $\Delta\varpi_{RMS}$ differences are localized to a pair of regions (100–120, 132–146) that are highlighted in grey. The secondary structure of the ground state of L99A T4L is illustrated. **b**, Values of $\Delta\varpi_{RMS}$ colour-coded onto the X-ray structure of L99A T4L (PDB: 3DMV[28]), ranging from blue ($\Delta\varpi_{RMS} = 0$) to red ($\Delta\varpi_{RMS} > 0.7$). The mesh surface indicates the position of the cavity formed by the Leu to Ala substitution at position 99. **c**, S^2 values for the backbone amide groups in the ground (blue) and excited (red) states of L99A T4L as predicted by the RCI approach[15]. **d**, Helix propensity values, predicted using TALOS+ (ref. 29), highlighting important changes in secondary structure between ground (blue) and excited (red) L99A T4L conformers.

state conformer. The lowest energy structures so obtained are in excellent agreement with the L99A crystal structure (r.m.s.d. of 0.6 ± 0.2 Å, Supplementary Table 6). Figure 2 shows an overlay of ten low energy representative excited state structures (Fig. 2a), along with the X-ray structure of the ground state (Fig. 2b) for comparison. As predicted on the basis of the input chemical shift data (Fig. 1a), there are clear structural differences between ground and excited states. These occur in a region immediately surrounding the cavity, involving rearrangement of the pair of short helices F and G that are orthogonal in the wild-type structure and that form a single, continuous and nearly straight helix in the excited state. This conformational rearrangement also includes a significant change in the backbone dihedral angle (Ψ) of Phe 114 to a helical value in the excited state ($+49°$ to $-36°$) and a reorientation of its side chain caused by a change in the torsion angle (χ_1) from a *gauche−* to a *trans* conformation (see below). The change in χ_1 projects the Phe 114 benzyl moiety into the cavity of L99A T4L, significantly decreasing its volume (Fig. 2c).

To cross-validate the excited state structures, we used Rosetta structure based design calculations to identify substitutions predicted to stabilize the excited state relative to the ground state (Supplementary Table 7). One such substitution is G113A, which replaces one of the most helix destabilizing residues (Gly) with the most favourable (Ala)[17] in a region of the structure that is predicted to become more helical in the excited state. ^1H-^{15}N and ^1H-^{13}C spectra of L99A,G113A T4L (recorded at low temperature (1 °C) to slow down the exchange and hence improve spectral quality) show two sets of cross-peaks that can be connected by magnetization exchange[18,19] (Fig. 3a, Supplementary Fig. 3). The first set corresponds to those observed for the L99A ground state, with a second set occurring at the positions predicted for the excited state on the basis of the chemical shifts obtained from CPMG relaxation dispersion experiments recorded on L99A T4L (Fig. 3b). Intensities of peaks from magnetization exchange experiments can be
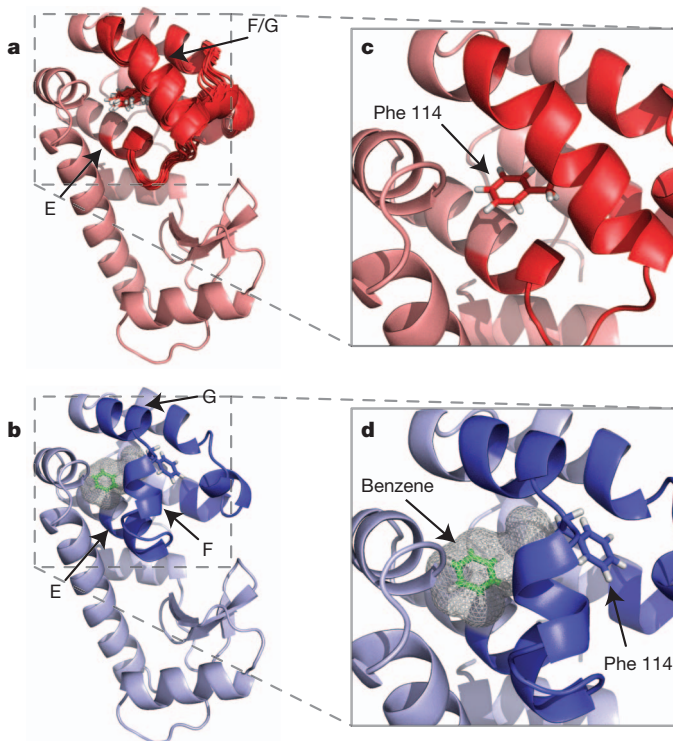
Figure 2 | The structure of the invisible, excited state of L99A T4L. **a**, Superposition of the 10 lowest energy structures of the L99A T4L excited state. The locations of helices E, F and G are indicated, along with the side chain of Phe 114 (see **c**). Only residues 100–120 and 132–146 were allowed to deviate from the L99A T4L ground state X-ray structure in calculations of the excited conformer (Methods). **b**, Ground state structure of L99A T4L, showing helices E, F and G and the position of Phe 114 (PDB: 3DMV[28]) or of benzene (green; PDB: 3DMX[28]) when it is bound inside the cavity (see **d**). **c**, **d**, Expanded regions of the excited state (**c**) and ground state (**d**) structures, focusing on the differences between helices F and G and the position of Phe 114.

fitted to extract the population of the excited state, p_E, and an exchange rate, $k_{ex} = k_{GE} + k_{EG}$ (k_{ij} is the exchange rate from state i to j); values of $p_E = 34 \pm 2\%$ and $k_{ex} = 48 \pm 1$ s^{-1}, at 1 °C, are obtained (Supplementary Fig. 4, Supplementary Table 8). Thus, the G113A mutation shifts the G $\rightleftharpoons$ E equilibrium, as expected from the excited state structure, from $p_E < 0.5\%$ to 34% at 1 °C.

In a previous set of studies based on analysis of ^{15}N and methyl ^{13}C CPMG relaxation dispersion profiles, we speculated that the excited state of L99A T4L is an open conformation where ligands can access the cavity[5]. The solution structure of the low populated L99A T4L conformer, however, predicts that hydrophobic ligands would not bind the excited state because the cavity is occupied by the side chain of Phe 114 (Fig. 2c). As a second cross-validation of the structure, we measured the binding of benzene to the ground and excited conformers independently using a sample of L99A,G113A T4L, where separate peaks can be observed for each state (Fig. 3a). A previous study has established that benzene binds to L99A T4L with a millimolar K_D and a dissociation rate of close to 1,000 s^{-1} at 20 °C (ref. 20). Lowering the temperature to 1 °C decreases both the rate of benzene binding and the rate of exchange between ground and excited states; these rates are reduced to the point where separate peaks are observed for the methyl group of Met 102 in ^1H-^{13}C HSQC spectra of the ground, excited and benzene-bound states of L99A, G113A T4L to which one molar equivalent of benzene was added (Fig. 3c). Rates of exchange between the three states can be quantified by analysis of magnetization exchange experiments[18,19,21]. From fits of the time-dependencies of the auto-peaks (labelled 'G', 'E', 'B' in Fig. 3c, diagonal panels of Fig. 3d) and cross-peaks (cross panels) to a model of three-site exchange (Fig. 3e and Supplementary
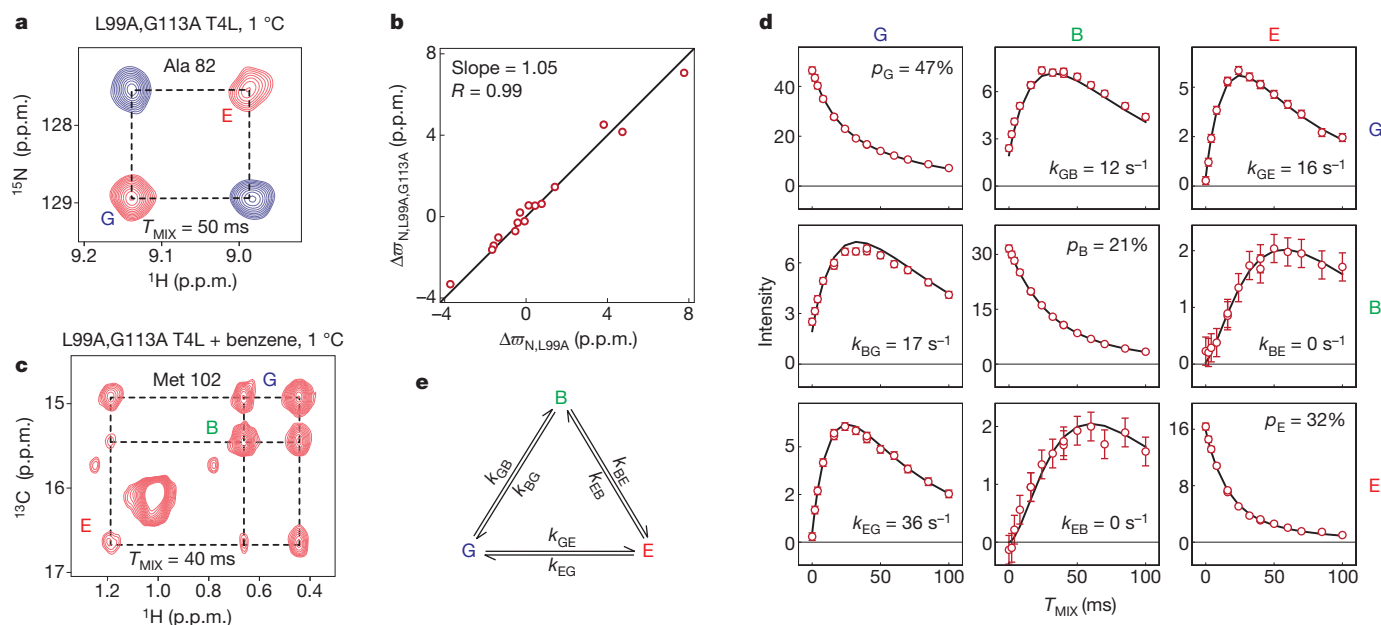
Figure 3 | Hydrophobic ligands do not bind the excited state of L99A T4L. a, Selected region of the ^1H–^{15}N correlation map from a magnetization exchange experiment recorded on L99A,G113A T4L at 1 °C, showing separate peaks for the ground (G) and excited (E) states. A pair of data sets are obtained, with the mixing time ($T_{MIX} = 50$ ms) recorded before or after the ^{15}N chemical shift evolution period, and the data sets subtracted so that diagonal- (cross-) peaks are positive (negative)[30]. **b**, Correlation between $\Delta\varpi_N$ values measured directly from the spectrum in **a** (y axis) and corresponding values from CPMG

relaxation dispersion measurements of L99A T4L (25 °C; x axis). **c**, Magnetization exchange spectrum ($T_{MIX} = 40$ ms) recorded on an L99A,G113A T4L sample with a 1:1 molar equivalent of benzene, focusing on Met 102 that shows well resolved correlations from ground, excited and benzene-bound (B) states. **d**, Intensity of auto- and cross-peaks for residue Met 102 from magnetization exchange experiments recorded as a function of T_{MIX} (red circles), along with the best fit of the data (solid lines) to the exchange model of **e**. Values of rates of exchange, k_{ij}, are indicated.

Table 8), the six relevant rates, k_{ij}, are extracted. Best fit values for k_{EB}, k_{BE} are $<0.1\,s^{-1}$ (Supplementary Fig. 5), and F-test analyses establish that there is no difference in the quality of the fits when these rates are set to zero, indicating that binding to the excited state does not occur. By contrast, $k_{GB} = 11.6 \pm 0.3\,s^{-1}$ and $k_{BG} = 17.4 \pm 0.4\,s^{-1}$, so that ligand binding proceeds via the ground state. The mechanism by which this occurs is not at present known, but it must involve excursions of the ground state to additional conformations, presumably on a timescale faster than those that are accessible to the dispersion experiments described here.

Whereas the G113A mutation shifts the fractional population of the excited state from approximately 3% to 34%, we were interested in perturbing the equilibrium still further. The R119P substitution is predicted by Rosetta to further favour the excited state because the X-ray structure of the L99A ground state[9] is incompatible with a Pro at position 119 due to steric clashes involving an H^δ proton and the C' of Thr 115. Consistent with this prediction, the ^1H–^{15}N HSQC spectrum of G113A,R119P,L99A T4L contains a single set of peaks (Supplementary Fig. 6) at resonance positions identical to those of the invisible, excited state of L99A that were determined by relaxation dispersion measurements (Supplementary Fig. 7). Further relaxation dispersion experiments recorded on the triple mutant established that the dominant conformer in solution, corresponding to the ligand inaccessible state (Supplementary Table 6), interconverts with a minor state conformer whose structure is that of the L99A ground state (Supplementary Fig. 8) (population of $3.8 \pm 0.1\%$, $k_{ex} = 806 \pm 28\,s^{-1}$, 1 °C). Thus, the pair of mutations G113A,R119P inverts the populations of ground and excited states relative to L99A T4L.

This population inversion, rendering the ligand-inaccessible state the major conformer, allows an additional test of the structure of the invisible L99A T4L excited state. Quantitative J-based scalar coupling experiments[22] recorded on L99A,G113A,R119P T4L confirm that the χ_1 rotamer state for Phe 114 is trans, as observed in the CS-Rosetta based structure of the excited state (Supplementary Fig. 9, Supplementary Table 9, see above). By contrast, similar experiments

recorded on the ground state of L99A T4L are consistent only with a gauche− conformation, as expected from the X-ray structure[9]. The trans χ_1 rotameric state for Phe 114 is a novel feature of the L99A T4L excited state and L99A,G113A,R119P T4L. In a G113A variant T4L, as examined here but where Leu is retained at position 99, a gauche− conformation is observed[23] for Phe 114; a trans χ_1 angle would lead to steric clashes with Leu 99 (Supplementary Fig. 10).

The L99A mutation in T4L creates an energy landscape in which a low-lying excited state is transiently populated. We have shown that this invisible, excited state has different functional properties from the
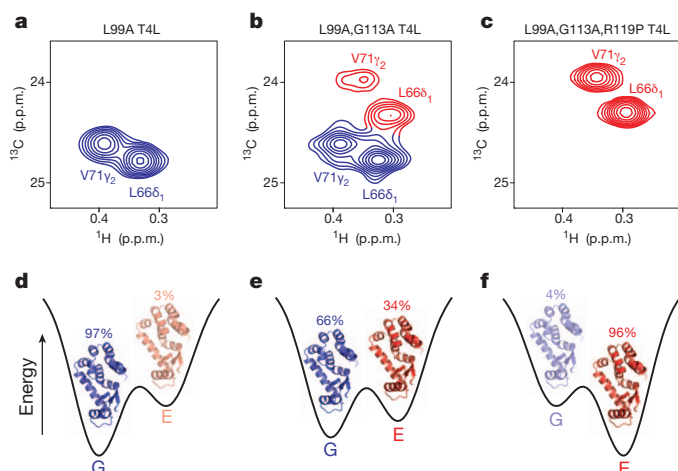
Figure 4 | The delicate balance between states on the energy landscape can be readily manipulated through mutation, providing a path for protein evolvability. a–c, Selected regions from ^1H–^{13}C HSQC spectra (recorded at 1 °C) of L99A T4L (**a**), L99A,G113A T4L (**b**) and L99A,G113A,R119P T4L (**c**), with the peaks from the ground and excited states coloured in blue and red, respectively. **d–f**, Corresponding energy landscapes, showing the structures of the ground and excited states and their fractional populations.

ground state—it does not bind hydrophobic ligands. This divergence in function can be controlled through mutation, with the G113A single mutation and the G113A,R119P double mutation changing the ratio of binding-competent to binding-incompetent states from 97% to 66% (G113A) and to less than 5% (G113A,R119P) (Fig. 4). The picture of evolving protein function suggested by our studies of L99A T4L is consistent with an emerging view of protein plasticity[24,25], with each molecule sampling a range of structures. Each unique conformer can, in turn, potentially carry out a different function[24,26,27]. A small number of mutations can then shift the relative populations of the conformers, thereby changing the activity of the protein, as has been observed in directed evolution experiments involving the introduction of mutations into flexible loop regions of enzymes[24]. Insight into the relation between protein dynamics, structure and evolvability is greatly facilitated through the powerful combination of relaxation dispersion NMR and Rosetta enabling individual protein states that populate the energy landscape to be investigated, even in cases where these conformers are invisible to other biophysical techniques. As applications of this methodology continue to grow, so too will our understanding of how protein dynamics control function, increasing the scope for the rational design of proteins with specific properties.

METHODS SUMMARY

Protein expression and purification. All NMR samples were prepared following previously published protocols, as described in detail in Methods.

NMR experiments and structure calculations. NMR experiments were recorded and analysed as described in Methods. Structure calculations of the L99A T4L excited state were based on experimental backbone ^{15}N, ^{13}C and ^1H chemical shifts obtained from CPMG dispersion experiments. The structures of regions in the excited state with significant variations in chemical shift values relative to those in the ground state were computed using the loop modelling application of Rosetta[16].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 March; accepted 30 June 2011.
Published online 21 August 2011.

1. Karplus, M. & Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl Acad. Sci. USA* **102**, 6679–6685 (2005).
2. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638–1642 (2006).
3. Henzler-Wildman, K. A. *et al.* Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838–844 (2007).
4. Eriksson, A. E., Baase, W. A., Wozniak, J. A. & Matthews, B. W. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* **355**, 371–373 (1992).
5. Mulder, F. A. A., Mittermaier, A., Hon, B., Dahlquist, F. W. & Kay, L. E. Studying excited states of protein by NMR spectroscopy. *Nature Struct. Biol.* **8**, 932–935 (2001).
6. Palmer, A. G., Kroenke, C. D. & Loria, J. P. NMR methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* **339**, 204–238 (2001).
7. Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl Acad. Sci. USA* **105**, 4685–4690 (2008).
8. Baase, W. A., Liu, L., Tronrud, D. E. & Matthews, B. W. Lessons from the lysozyme of phage T4. *Protein Sci.* **19**, 631–641 (2010).
9. Eriksson, A. E. *et al.* Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255**, 178–183 (1992).
10. Korzhnev, D. M. & Kay, L. E. Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc. Chem. Res.* **41**, 442–451 (2008).
11. Hansen, D. F., Vallurupalli, P. & Kay, L. E. Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *J. Biomol. NMR* **41**, 113–120 (2008).
12. Cavalli, A., Salvatella, X., Dobson, C. M. & Vendruscolo, M. Protein structure determination from NMR chemical shifts. *Proc. Natl Acad. Sci. USA* **104**, 9615–9620 (2007).
13. Vallurupalli, P., Hansen, D. F. & Kay, L. E. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc. Natl Acad. Sci. USA* **105**, 11766–11771 (2008).
14. Korzhnev, D. M., Religa, T. L., Banachewicz, W., Fersht, A. R. & Kay, L. E. A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329**, 1312–1316 (2010).
15. Berjanskii, M. V. & Wishart, D. S. A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* **127**, 14970–14971 (2005).
16. Wang, C., Bradley, P. & Baker, D. Protein-protein docking with backbone flexibility. *J. Mol. Biol.* **373**, 503–519 (2007).
17. Fersht, A. *Structure and Mechanism in Protein Science* (Freeman and Company, 1999).
18. Montelione, G. T. & Wagner, G. 2D chemical-exchange NMR-spectroscopy by proton detected heteronuclear correlation. *J. Am. Chem. Soc.* **111**, 3096–3098 (1989).
19. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR* **4**, 727–734 (1994).
20. Feher, V. A., Baldwin, E. P. & Dahlquist, F. W. Access of ligands to cavities within the core of a protein is rapid. *Nature Struct. Biol.* **3**, 516–521 (1996).
21. Religa, T. L., Sprangers, R. & Kay, L. E. Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science* **328**, 98–102 (2010).
22. Hu, J. S., Grzesiek, S. & Bax, A. Two-dimensional NMR methods for determining χ_1 angles of aromatic residues in proteins from three-bond $J_{C'C\gamma}$ and $J_{NC\gamma}$ couplings. *J. Am. Chem. Soc.* **119**, 1803–1804 (1997).
23. Nicholson, H., Tronrud, D. E., Becktel, W. J. & Matthews, B. W. Analysis of the effectiveness of proline substitutions and glycine replacements in increasing the stability of phage T4 lysozyme. *Biopolymers* **32**, 1431–1441 (1992).
24. Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).
25. Jensen, R. A. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**, 409–425 (1976).
26. Walden, W. E. *et al.* Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. *Science* **314**, 1903–1908 (2006).
27. Tang, C., Schwieters, C. D. & Clore, G. M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449**, 1078–1082 (2007).
28. Liu, L. J., Baase, W. A. & Matthews, B. W. Halogenated benzenes bound within a non-polar cavity in T4 lysozyme provide examples of I⋯S and I⋯Se halogen-bonding. *J. Mol. Biol.* **385**, 595–605 (2009).
29. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).
30. Rodríguez, J. C., Jennings, P. A. & Melacini, G. Using chemical exchange to assign non-covalent protein complexes in slow exchange with the free state: enhanced resolution and efficient signal editing. *J. Biomol. NMR* **30**, 155–161 (2004).

METHODS

Protein expression and purification. The gene for expression of L99A T4L was optimized for protein production (Genscript) and placed in a pET-29b plasmid. Additional plasmids for the expression of the mutants used in this study (L99A,G113A T4L; L99A,G113A,R119P T4L) were constructed from the L99A T4L plasmid. T4L proteins were expressed in *Escherichia coli* BL21(DE3) cells grown in M9 minimal media with glucose (3–4 g l^{-1}, typically) and ammonium chloride (NH$_4$Cl, 1 g l^{-1}) as the sole carbon and nitrogen sources, respectively. Proteins with specific labelling patterns (see below) were obtained by expression in M9 media containing the appropriately labelled glucose, ^{15}NH$_4$Cl and solvent (H$_2$O or D$_2$O)[31]. The labelling patterns (carbon source/solvent) used in this study are: uniform (U) ^{15}N (glucose/H$_2$O), U-^{13}C/^{15}N ([^{13}C$_6$]-glucose/H$_2$O), U-^{2}H/^{13}C/^{15}N ([^{13}C$_6$,^{2}H$_7$]-glucose/100% D$_2$O), U-^{2}H/^{15}N ([^{2}H$_7$]-glucose/100% D$_2$O), ^{13}C$^\alpha$/^{15}N ([2-^{13}C]-glucose/100% H$_2$O), U-^{13}C/U-^{15}N/50%^{2}H ([^{13}C$_6$,^{2}H$_7$]-glucose/50% D$_2$O), ^{13}CH$_3$-Met/^{15}N (glucose/H$_2$O, supplemented with 100 mg l^{-1} of ^{13}CH$_3$-Met added to the media 30 min before induction of protein overexpression). *E. coli* BL21(DE3) cells transformed with the appropriate plasmid were grown in one or two litres of media at 37 °C until an OD$_{600}$ of ∼1. The temperature was then reduced to 16 °C and protein expression was induced with 1 mM IPTG for 14–18 h. The cells were harvested by centrifugation and frozen. The protein was purified from the cells as described[32].

Samples. NMR samples (∼1.5 mM in protein) were prepared in a buffer consisting of 50 mM sodium phosphate, 25 mM NaCl, 2 mM EDTA, 2 mM NaN$_3$, pH 5.5 in either 10% or 100% D$_2$O. The experiments were performed on Varian Inova spectrometers operating at frequencies (^{1}H) of 500, 600 and 800 MHz, at a temperature of 25 °C, unless stated otherwise.

Assignments. Complete assignments for L99A T4L have been reported previously[5]. The major (ground) state peaks in the L99A,G113A T4L mutant were assigned by comparison with assigned spectra of L99A T4L. Minor (excited-) state ^{1}H–^{15}N resonance assignments were obtained using an ^{15}N magnetization exchange experiment[19] recorded at 800 MHz, 1 °C, with a mixing time (T_{MIX}) of 50 ms. At 1 °C the excited state is populated to ∼34%, k_{ex} ≈ 50 s^{-1}, so that very clear exchange peaks correlate ground and excited state resonances. Backbone ^{1}H/^{13}C/^{15}N and ^{13}C$^\beta$ assignments for the L99A,G113A,R119P T4L mutant were obtained using standard triple resonance experiments[33] recorded at 34 °C either at 500 or 800 MHz. Assignments were very close to complete. The elevated temperature (34 °C) and lower field (500 MHz) were used to minimize signal loss due to chemical exchange.

CPMG experiments. ^{1}H/^{15}N/^{13}C constant-time (CT) CPMG relaxation dispersion experiments were (usually) performed at two static magnetic fields. Typically, dispersion curves were composed of a large number (∼15) of ν_{CPMG} values with errors estimated based on two or three repeat values[34]. Here $\nu_{CPMG} = 1/(4\tau_{CPMG})$, where $2\tau_{CPMG}$ is the spacing between the refocusing π pulses applied during the CT delay of length T_{Relax}. Details of the experiments[6,35,36] used to characterize the excited state of L99A T4L are summarized in Supplementary Table 1.

Sign experiments. Single-quantum CPMG relaxation dispersion experiments provide only the magnitude of the change in chemical shift $|\Delta\varpi_{GE}| = |\varpi_E - \varpi_G|$ between the two exchanging states. Signs were obtained by comparing the positions of the ground state peaks in HSQC spectra recorded at different static magnetic fields and/or between peak positions in HSQC/HMQC spectra recorded at the same field[37,38]. Once the signs of amide nitrogen $\Delta\varpi$ values were obtained, the corresponding signs of amide protons were generated from zero-quantum (ZQ) and double-quantum (DQ) CPMG experiments[39]. Experiments used to obtain this information are listed in Supplementary Table 2.

Quantitative J-modulated experiments. χ_1 angles of aromatic residues in both L99A T4L (ground state) and L99A,G113A,R119P T4L (ground state that is a mimic of the L99A T4L excited state structure, see Supplementary Fig. 7 and text) were determined by measurement of three-bond $J_{C'C\gamma}$ and $J_{NC\gamma}$ scalar couplings using quantitative J-based experiments[22,40] recorded with dephasing delays of 100 ms ($J_{C'C\gamma}$) and 120 ms ($J_{NC\gamma}$), respectively. All experiments were obtained at 35 °C on a 600 MHz spectrometer, using uniformly ^{15}N, ^{13}C enriched samples. Difference spectra for both experiments and both proteins are shown in Supplementary Fig. 9. Measured scalar coupling values are summarized in Supplementary Table 9 along with the χ_1 angles for the aromatic residues of the ground and excited state structures.

Magnetization-exchange experiments. ^{15}N magnetization exchange experiments, recorded at 1 °C, 800 MHz, were used for assignment of excited state correlations of L99A,G113A T4L, as described above. A pair of experiments were recorded in which the exchange mixing period, T_{MIX} (50 ms), was placed (1) before and (2) after indirect detection of ^{15}N magnetization. Subtraction of the two data sets so obtained generates a two-dimensional spectrum where correlations from ground and excited states (positive) are connected by cross-peaks (negative), forming a 'rectangular' structure[30] (Fig. 3a). Quantitative methyl ^{13}C

magnetization exchange experiments to quantify exchange, $G \underset{k_{EG}}{\overset{k_{GE}}{\rightleftharpoons}} E$ (see below or text), in L99A,G113A T4L were performed at 600 MHz, 1 °C, using a sample in which only Met-C$^\varepsilon$ was ^{13}C enriched. A second similarly labelled sample to which a small amount of benzene was added (approximately 1:1 molar equivalents of benzene and protein) was used to study exchange between the ground, excited and benzene-bound states of L99A,G113A T4L. Experiments were recorded with T_{MIX} values ranging between 0 and 85 ms (100 ms in the presence of benzene) with errors estimated based on repeat measurements.

Data processing. The NMRpipe software package[41] was used to processes all of the NMR data. Subsequent visualization and peak picking was achieved using the program Sparky[42]. The intensities of peaks (I) were obtained using the program FuDA (http://pound.med.utoronto.ca/software.html), while the CcpNmr set of programs[43] was used to analyse some of the triple resonance assignment experiments.

Analysis of CPMG data. Relaxation dispersion (RD) profiles, $R_{2,eff}(\nu_{CPMG})$, were generated from peak intensities, $I(\nu_{CPMG})$, measured in a series of 2D correlation maps recorded at various CPMG frequencies, ν_{CPMG}. The effective relaxation rates, $R_{2,eff}(\nu_{CPMG})$, were computed via the relation:

$$R_{2,eff}(\nu_{CPMG}) = -\frac{1}{T_{Relax}} \ln\left(\frac{I(\nu_{CPMG})}{I_o}\right)$$

where I_o is the peak intensity extracted from a reference spectrum recorded without the CPMG block. RD profiles were analysed assuming a two-site exchanging system, $G \rightleftharpoons E$, where the major state, G, interconverts with the minor state, E, as described previously in the context of the L99A system[5,44]. The model parameters defining the chemical exchange process, that is the exchange rate, k_{ex}, the population of the minor state, p_E, and the absolute difference in chemical shifts between the two states, $|\Delta\omega_{GE}| = |\omega_E - \omega_G|$, were determined by minimizing the target function:

$$\chi^2(\zeta) = \sum_{i=1}^{N} \left(\frac{R_{2,eff}^{Exp} - R_{2,eff}^{Calc}(\zeta)}{\Delta R_{2,eff}^{Exp}}\right)^2$$

where $R_{2,eff}^{Exp}$ and $\Delta R_{2,eff}^{Exp}$ are the experimental effective transverse relaxation rates and their associated uncertainties, $R_{2,eff}^{Calc}(\zeta)$ are back-calculated relaxation rates obtained by numerical integration of the Bloch-McConnell equations[45] using the program CATIA (http://pound.med.utoronto.ca/software.html), ζ represents the set of adjustable model parameters and the sum is over all the experimental data points.

Analysis of quantitative magnetization exchange data. As described above and in the text, methyl ^{13}C magnetization exchange experiments were recorded on L99A,G113A T4L (1 °C) without and with added benzene using a pulse scheme described previously[21,46]. Data from Met 102 were analysed because separate, well-resolved correlations are obtained for the ground, excited and benzene-bound states that could be accurately quantified. At the low temperature used (1 °C), the interconversion between ground and excited states as well as benzene binding are in the slow exchange regime, a requirement for the magnetization exchange experiment. The intensity I of auto- (cross-) peaks corresponding to magnetization that is not (is) transferred between states during a mixing time, T_{MIX}, has been analysed by numerically solving the Bloch-McConnell equations for the evolution of magnetization in the presence of chemical exchange[45]. The time dependence of magnetization during the entire pulse sequence was simulated (details are available from the authors on request). The resonance frequencies of the peaks and their transverse relaxation rates were obtained from the positions and linewidths of the peaks in spectra, respectively. The fitted parameters include the total magnetization, longitudinal relaxation rates (R_1) for each state, the fractional populations of each state, p_i (subject to the constraint that $\sum_i p_i = 1$), and the rates of exchange between states i and j, $k_{exij} = k_{ij} + k_{ji}$. The fitting parameters were optimized using a simplex procedure to minimize the function:

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{I_i^{Exp} - I_i^{Calc}}{\sigma_i^{Exp}}\right)^2$$

Here the summation is over all the experimental data points, I^{Exp} is the experimental intensity, I^{Calc} is the calculated intensity and σ^{Exp} is the error in the intensity. In the case of two-state exchange (no benzene), 48 data points from four peaks associated with Met 102 (two auto- and two cross- peaks × 12 T_{MIX} values) were fitted using five fitting parameters. For the three-state exchange (approximately 1:1 molar equivalent protein:benzene), 117 data points from nine Met 102 peaks (three auto- and six cross-peaks × 13 T_{MIX}; Fig. 3d) were fitted using nine fitting parameters. The minimum error in the intensities was assumed to be 3%. Errors in the fitted parameters were estimated using a Monte Carlo

procedure[47]. Here 50 synthetic data sets were generated using the best-fit parameters in which random error was added to magnetization intensities and $^{13}C/^1H$ R_2 values (based on the experimental errors) and each of the data sets fitted as per the experimental data. Errors are calculated as 1 s.d. in the extracted values.

These experiments clearly indicate that only the ground state binds benzene. As shown in Fig. 3d, fits of the time dependencies of diagonal- and cross-peaks from magnetization exchange spectra establish that $k_{BE}, k_{EB} < 0.1\,s^{-1}$. To further support the result that benzene does not bind the excited state, Supplementary Fig. 5 plots the reduced χ^2 obtained from the fit of the magnetization exchange data as a function of $k_{exBE} = k_{BE} + k_{EB}$. A clear minimum occurs for $k_{exBE} \approx 0$ ($\chi^2_{red} = 1.1$). Of note, when k_{exBE} is fixed to the relatively small value of $0.5\,s^{-1}$, χ^2_{red} increases by fivefold (to 5.4), clearly indicating that k_{exBE} is very small. From the principle of detailed balance for the equilibrium denoted in Fig. 3e, it is predicted that the ratios $\dfrac{p_E}{p_E + p_G}, \dfrac{k_{GE}}{k_{EG}}$ will be independent of ligand (benzene) concentration, as observed to within experimental error (Supplementary Table 8). As expected, binding of benzene shifts the equilibrium from the excited (binding incompetent) to the ground/bound states. These results are in complete agreement with the structure of the excited state of L99A T4L, showing clearly that Phe 114 is inserted into the cavity, hence obstructing the binding of hydrophobic ligands.

L99A excited state chemical shifts. Excited state amide nitrogen, amide proton and carbonyl chemical shifts are available for nearly all the (164) L99A T4L residues (Supplementary Tables 3–5). In cases where the sign of $\Delta\varpi_{GE}$ is obtained, the chemical shift of the excited state is readily calculated, $\varpi_E = \varpi_G + \Delta\varpi_{GE}$. In cases where the sign is not available, the excited state chemical shift is ambiguous but restricted to two values, $\varpi_E = \varpi_G + \Delta\varpi_{GE}$ or $\varpi_E = \varpi_G - \Delta\varpi_{GE}$. The magnitude of the change in chemical shift, $|\Delta\varpi_{GE}|$, also provides useful information (see section on Rosetta calculations below).

Calculation of L99A T4L excited state structures. The structure of the excited state of L99A T4L was obtained by using the CS-Rosetta approach developed for the determination of ground state protein structures[7], with a number of important differences relative to the standard protocol. Based on the $\Delta\varpi_{RMS}$ values (see text), we assumed that only the regions encompassing residues 100–120 and 132–146 adopt different conformations in the ground and excited states. The structure of the rest of the molecule was fixed to the ground state crystal structure of L99A T4L (3DMV[28]). The structures of these two regions in the excited state were computed using the loop modelling application of Rosetta[16]. As a control, identical CS-Rosetta computations of the ground state structure were also performed using the same limited set of ground state shifts (that is, that are available for the excited state). To avoid any bias, T4 lysozyme structures were removed from the fragment databases in all the computations described here. Two hundred starting 3mer and 9mer fragments were selected for each position using the CS-Rosetta approach[7], modified to include ambiguous excited state shifts in cases where the sign of $\Delta\varpi_{GE}$ could not be determined. Fragments were scored against ambiguous shifts by selecting the shift which agreed best with the one predicted for the fragment. Ambiguous chemical shifts were similarly taken into account during the scoring of the final structures. The selected fragments were then used in a standard Rosetta loop modelling protocol[16] to generate 9,600 structures (the target secondary structure propensities input into Rosetta are those predicted by TALOS+[29]). Supplementary Fig. 2a and b plots the energies of the resultant ground state structures, generated with ground state chemical shifts (Supplementary Fig. 2a, CS-Rosetta energy; Supplementary Fig. 2b, chemical shift component of the CS-Rosetta energy term) versus r.m.s.d. to the L99A ground state X-ray structure. The characteristic funnel shape energy profile so obtained is an excellent indicator of convergence and indeed the lowest energy 10 (50) structures have pair-wise backbone r.m.s.d. values of 0.6 ± 0.15 (0.55 ± 0.15) Å relative to the ground state crystal structure, including only those residues that were allowed to move in the calculations. The corresponding plots for the excited state structures are shown in Supplementary Fig. 2c and d.

Some of the low energy CS-Rosetta structures produced with the excited state chemical shifts are very similar to the ground state structure (Supplementary Fig. 2c). This is not surprising, given the experimental finding that the ground state is more stable than the excited state by $\sim2\,kcal\,mol^{-1}$. However, the chemical shift score by itself clearly indicates that the ground state structure is not a 'good' solution (Supplementary Fig. 2d). Hence a two-step selection procedure was used. Out of the 9,600 structures that were generated initially, 960 structures with the best CS-Rosetta score were selected for further analysis. As a second selection step, 96 of these structures with the best chemical shift score were retained. In this way, structures with both low energy CS-Rosetta and low energy chemical shift scores are selected. It is noteworthy that the structures so generated were essentially identical irrespective of whether selection was first performed on the basis of the Rosetta score followed by selection according to chemical shifts (as described here) or whether the opposite order of scoring was used. We find that there is a major cluster of structures with an r.m.s.d. of ~1.4 Å to the ground state crystal

structure and a much smaller cluster with an r.m.s.d. of ~1 Å to the ground state conformer; the final 96 structures selected are indicated in green in Supplementary Fig. 2c and d. Only one out of the forty lowest energy structures after the second round of filtering is part of the second cluster (10% of the 96 structures). A major difference between the two sets of structures is the χ_1 angle of Phe 114, with this dihedral angle assuming a *trans* (*gauche*−) conformation in the major (minor) cluster. Quantitative-*J* experiments[22,40] recorded on the L99A,G113A,R119P T4L mutant that is an excellent mimic of the excited state (Supplementary Fig. 7, Supplementary Table 6, see text) clearly show that the *trans* conformation is the only one populated (Supplementary Fig. 9, Supplementary Table 9). Notably, CS-Rosetta computations performed with the full set of chemical shifts for this mutant (L99A,G113A,R119P T4L) produced structures that have a conformation that is essentially identical to that obtained for the excited state of L99A T4L (Supplementary Table 6). All structure calculations were performed on the University of Toronto SciNet super-computer cluster[48]. Each set of 9,600 structures required 10 h of computational time using 256 processor cores. Pymol[49] and Chimera[50] were used to visualize and analyse the resultant structures.

Mutations shifting ground and excited state populations. Predictions of free energy differences ($\Delta\Delta G$) were performed as described[51]. Briefly, single point mutants of the 19 amino acids (except cysteine) were made *in silico* in the region corresponding to residues 105–120 and $\Delta\Delta G$ values were computed for representative structures of both ground and excited states. The crystal structure of L99A T4L (3DMV[28]) was used as the ground state, with a representative low energy structure obtained from CS-Rosetta simulations performed with excited state chemical shifts used for the excited state. The free energy difference between corresponding point mutants in the excited and ground states was computed ($\Delta\Delta G_E - \Delta\Delta G_G$; negative values indicate relative stabilization of the excited state). We screened for single point mutants that energetically favoured the excited state conformation and simultaneously disfavoured the ground state conformation (Supplementary Table 7). Mutations for experimental characterization were selected according to two criteria: secondary structure propensity of the excited state and Rosetta $\Delta\Delta G$ predictions.

31. Lundström, P., Vallurupalli, P., Hansen, D. F. & Kay, L. E. Isotope labeling methods for studies of excited protein states by relaxation dispersion NMR spectroscopy. *Nature Protocols* **4,** 1641–1648 (2009).
32. Vallurupalli, P., Hansen, D. F., Lundstrom, P. & Kay, L. E. CPMG relaxation dispersion NMR experiments measuring glycine $^1H\alpha$ and $^{13}C\alpha$ chemical shifts in the 'invisible' excited states of proteins. *J. Biomol. NMR* **45,** 45–55 (2009).
33. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34,** 93–158 (1999).
34. Korzhnev, D. M. *et al.* Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430,** 586–590 (2004).
35. Vallurupalli, P., Hansen, D. F., Stollar, E., Meirovitch, E. & Kay, L. E. Measurement of bond vector orientations in invisible excited states of proteins. *Proc. Natl Acad. Sci. USA* **104,** 18473–18477 (2007).
36. Hansen, D. F., Vallurupalli, P., Lundstrom, P., Neudecker, P. & Kay, L. E. Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: how well can we do? *J. Am. Chem. Soc.* **130,** 2667–2675 (2008).
37. Skrynnikov, N. R., Dahlquist, F. W. & Kay, L. E. Reconstructing NMR spectra of "invisible" excited protein states using HSQC and HMQC experiments. *J. Am. Chem. Soc.* **124,** 12352–12360 (2002).
38. Bouvignies, G. *et al.* A simple method for measuring signs of $^1H^N$ chemical shift differences between ground and excited protein states. *J. Biomol. NMR* **47,** 135–141 (2010).
39. Orekhov, V. Y., Korzhnev, D. M. & Kay, L. E. Double- and zero-quantum NMR relaxation dispersion experiments sampling millisecond time scale dynamics in proteins. *J. Am. Chem. Soc.* **126,** 1886–1891 (2004).
40. Hu, J. S. & Bax, A. Determination of φ and χ_1 angles in proteins from ^{13}C-^{13}C three-bond J couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond? *J. Am. Chem. Soc.* **119,** 6360–6368 (1997).
41. Delaglio, F. *et al.* NMRpipe – a multidimensional spectral processing system based on Unix pipes. *J. Biomol. NMR* **6,** 277–293 (1995).
42. Goddard, T. D. & Kneller, D. G. SPARKY 3 (University of California, San Francisco, 2006).
43. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins Struct. Funct. Bioinform.* **59,** 687–696 (2005).
44. Mulder, F. A., Hon, B., Muhandiram, D. R., Dahlquist, F. W. & Kay, L. E. Flexibility and ligand exchange in a buried cavity mutant of T4 lysozyme studied by multinuclear NMR. *Biochemistry* **39,** 12614–12622 (2000).
45. McConnell, H. M. Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* **28,** 430–431 (1958).
46. Sprangers, R., Gribun, A., Hwang, P. M., Houry, W. A. & Kay, L. E. Quantitative NMR spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release. *Proc. Natl Acad. Sci. USA* **102,** 16678–16683 (2005).
47. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes in C. The Art of Scientific Computing* 2nd edn (Cambridge Univ. Press, 1992).
48. Loken, C. *et al.* SciNet: Lessons learned from building a power-efficient top-20 system and data centre. *J. Phys. Conf. Ser.* **256,** 012026 (2010).

49. The PyMOL Molecular Graphics System, Version 1.3 (Schrödinger, LLC, 2010).
50. Pettersen, E. F. *et al.* UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
51. Kellogg, E. H., Leaver-Fay, A. & Baker, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **79,** 830–838 (2011).

LETTER

Protection of repetitive DNA borders from self–induced meiotic instability

Gerben Vader[1]*, Hannah G. Blitzblau[1]*, Mihoko A. Tame[1], Jill E. Falk[1]†, Lisa Curtin[1,2] & Andreas Hochwagen[1]†

DNA double strand breaks (DSBs) in repetitive sequences are a potent source of genomic instability, owing to the possibility of non-allelic homologous recombination (NAHR). Repetitive sequences are especially at risk during meiosis, when numerous programmed DSBs are introduced into the genome to initiate meiotic recombination[1]. In the repetitive ribosomal DNA (rDNA) array of the budding yeast *Saccharomyces cerevisiae*, meiotic DSB formation is prevented in part through Sir2-dependent heterochromatin formation[2,3]. Here we show that the edges of the rDNA array are exceptionally susceptible to meiotic DSBs, revealing an inherent heterogeneity in the rDNA array. We find that this localized DSB susceptibility necessitates a border-specific protection system consisting of the meiotic ATPase Pch2 and the origin recognition complex subunit Orc1. Upon disruption of these factors, DSB formation and recombination increased specifically in the outermost rDNA repeats, leading to NAHR and rDNA instability. Notably, the Sir2-dependent heterochromatin of the rDNA itself was responsible for the induction of DSBs at the rDNA borders in *pch2Δ* cells. Thus, although the activity of Sir2 globally prevents meiotic DSBs in the rDNA, it creates a highly permissive environment for DSB formation at the junctions between heterochromatin and euchromatin. Heterochromatinized repetitive DNA arrays are abundant in most eukaryotic genomes. Our data define the borders of such chromatin domains as distinct high-risk regions for meiotic NAHR, the protection of which may be a universal requirement to prevent meiotic genome rearrangements that are associated with genomic diseases and birth defects.

To understand better the mechanisms that protect repetitive DNA from meiotic NAHR, we analysed the single tandem rDNA array of budding yeast. Meiotic DSB formation and recombination in the rDNA are repressed by the histone deacetylase Sir2 (refs 2, 3). Additionally, Pch2, a widely conserved meiosis-specific ATPase, suppresses meiotic recombination in the rDNA by an unknown mechanism[4,5]. We used clamped-homogenous electric field (CHEF) electrophoresis and Southern blotting of excised rDNA arrays to address whether Pch2 regulates meiotic DSB formation in the rDNA. Consistent with previous results[2,3], the level of full-length rDNA arrays that remained 8 h after meiotic induction was significantly reduced in *sir2Δ* mutants compared to wild-type cells, indicating increased DSB formation (Fig. 1a and Supplementary Fig. 1a). By contrast, no such reduction occurred in *pch2Δ* mutants, although we observed a tenfold increase in crossover recombination across the rDNA array (Fig. 1a, b). Because small changes in array length would not be detectable by the CHEF gel assay, we wondered whether DSB formation in *pch2Δ* mutants occurred specifically in the outermost rDNA repeats. To test this possibility, we generated *pch2Δ* strains carrying a *URA3* insertion at defined positions in the rDNA array (Fig. 1c) and analysed the rDNA repeat units that directly flank these insertions by Southern blotting. We observed a strongly DSB-prone site in repeat 1 and weak DSB

formation in repeat 3, whereas no DSB formation was detectable in repeat 10 of the approximately 100 rDNA repeats (Fig. 1d). Thus, *pch2Δ* cells undergo increased meiotic DSB formation predominantly in the outermost rDNA repeats.

To determine whether *PCH2* suppresses DSB formation only within the rDNA, or also in other regions of the genome, we first analysed a chromosomal fragment spanning the junction between single-copy DNA and rDNA in a *pch2Δ* mutant by Southern blotting. We observed additional, strong DSB formation in the adjacent single-copy sequences (Fig. 1e and Supplementary Fig. 1b), which were previously shown to have exceptionally low levels of meiotic DSBs in *PCH2* cells[6,7] (Fig. 1f). The observed break sites behaved similarly to known meiotic DSBs[8]; they were induced during meiosis in *dmc1Δ* and *DMC1* cells (Fig. 1d, e and Supplementary Fig. 1c), depended on the meiotic DSB machinery (Supplementary Fig. 1d)[9], promoted meiotic recombination (Supplementary Fig. 1e) and occurred in gene promoters (Fig. 1e and Supplementary Fig. 1b). Indeed, even the DSBs observed in repeat 1 mapped to the promoter of a gene (*TAR1*) that is encoded in every rDNA repeat[10] (Fig. 1e). Genome-wide analysis of DSBs[6] in *pch2Δ* cells showed that strong DSB induction occurred in 30–50-kilobase (kb) regions of single-copy sequence abutting both sides of the rDNA (Fig. 1f). Increased DSB formation was also observed close to other heterochromatic regions (telomeres and *HML*), whereas the DSB landscape elsewhere in the genome was not markedly altered (Supplementary Figs 1f, g, 2 and Supplementary Table 1). In contrast to *pch2Δ* mutants, the loss of *SIR2* did not lead to increased DSB formation adjacent to the rDNA array (Fig. 1f). Thus, Pch2 represses recombination in the rDNA at the level of DSB formation, but in a manner distinct from Sir2.

We investigated whether the increased DSB formation in the outermost rDNA repeats in *pch2Δ* mutants (Fig. 1d) resulted in a local increase in rDNA recombination. We measured recombination rates using flanking markers to the left and right of the rDNA together with a collection of single *URA3* insertions tiling inwards from the left side of the rDNA (Fig. 1c). Analysis of a *URA3* insertion in the centre of the rDNA (inserted next to repeat 49 of 99) indicated that recombination occurred in a symmetrical pattern. Notably, about 80% of the recombination events in the left half of the rDNA occurred within the first ten repeats from the left border (Fig. 1g and Supplementary Table 2), with about 30% taking place within repeat 1. Thus, there is a strong bias for recombination in the rDNA repeats very close to the array border.

Because recombination in repetitive DNA can lead to NAHR, we selected tetrads of *pch2Δ* mutants that had undergone recombination in the rDNA, and determined the resulting rDNA repeat number between the *URA3* insertion and the left rDNA boundary. In 70% (*n* = 47) of the tetrads investigated from different *URA3* integrants, we detected changes in repeat number, ranging from 1 to 19 repeats (Fig. 1i, j and Supplementary Table 2), demonstrating that rDNA crossovers in *pch2Δ* cells are frequently associated with NAHR.

[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2]Somerville High School, Somerville, Massachusetts 02143, USA. †Present addresses: David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA (J.E.F.); Department of Biology, New York University, 100 Washington Square East, New York, New York 10003, USA (A.H.).
*These authors contributed equally to this work.
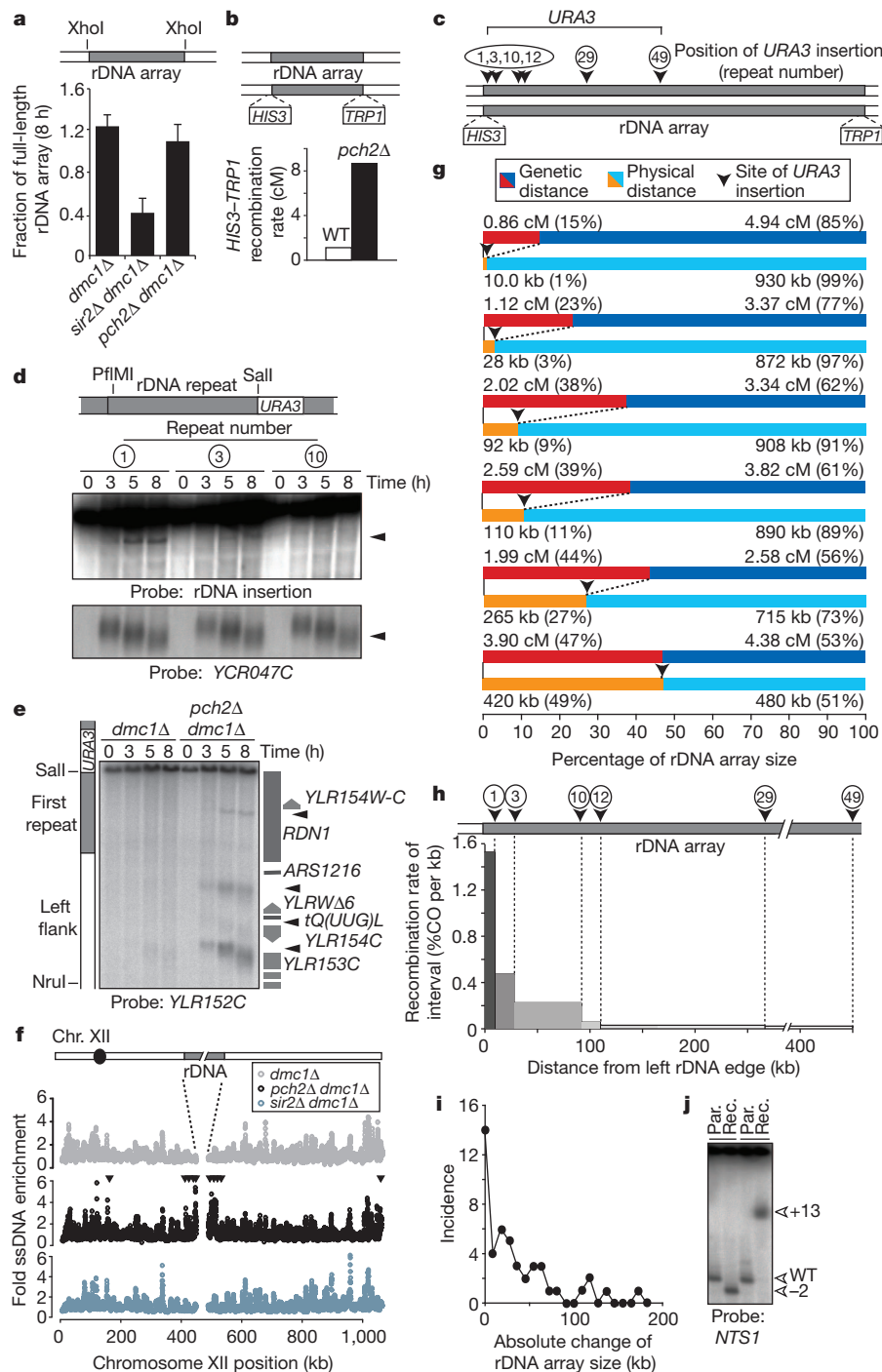
Figure 1 | Ribosomal-DNA-associated DSB formation and recombination.
a, CHEF analysis of the rDNA of meiotic *dmc1Δ* (H5217), *sir2Δ dmc1Δ* (H2953) and *pch2Δ dmc1Δ* (H5216) cells. The schematic shows the analysed XhoI restriction fragment. A *dmc1Δ* mutation was used to prevent DSB repair. The mean (+ s.e.m.) of five experiments is shown. Significance was assessed by one-tailed Student's *t*-test: *dmc1Δ* versus *sir2Δ dmc1Δ*, *P* value = 0.00122; *dmc1Δ* versus *pch2Δ dmc1Δ*, *P* value = 0.254; *pch2Δ dmc1Δ* versus *sir2Δ dmc1Δ*, *P* value = 0.00216. **b**, Schematic of markers inserted in unique single-copy sequences within 500 base pairs (bp) of the rDNA, and crossover rates in wild-type (WT; H3026; *n* = 467) and *pch2Δ* (H3027; *n* = 186) cells. **c**, Schematic indicating marker locations in the rDNA used in **d** and **g**. *URA3* markers were inserted in the *NTS1/2* region of the indicated repeats. **d**, Southern blot for restriction fragments containing the indicated insert-associated rDNA repeat units from *pch2Δ dmc1Δ* strains H5622 (repeat 1), H5636 (repeat 3) and H5706 (repeat 10). DNA was digested with PflMI and SalI and probed for the unique rDNA insertion. The *YCR047C* probe (on DNA digested with HindIII) was a positive control for DSB formation. **e**, Southern blot of the left rDNA flank,

including the outermost rDNA repeat, in *dmc1Δ* (H5583) and *pch2Δ dmc1Δ* (H5622) cells (SalI and NruI digest; probe *YLR152C*). Positions of open reading frames are shown schematically alongside the Southern blot. **f**, Enrichment profile of ssDNA in chromosome XII in *dmc1Δ* (H118, light grey), *pch2Δ dmc1Δ* (H2629, black) and *sir2Δ dmc1Δ* (H2953, dark grey) cells. Arrowheads indicate >twofold increased DSB formation in *pch2Δ dmc1Δ* compared to *dmc1Δ* cells. **g**, Tetrad analysis of *pch2Δ URA3*-rDNA-insertion strains H4611 (repeat 1), H4613 (repeat 3), H3823 (repeat 10), H4612 (repeat 12), H3820 (repeat 29) and H3821 (repeat 49; Supplementary Table 2). Recombination rates between *URA3* and rDNA-flanking markers are shown in relation to the physical *URA3* positions in the rDNA. **h**, Relative contribution of each measured interval indicated in **g** to total rDNA recombination (percentage of crossovers (CO) per kb of interval). **i**, Incidence of changes in rDNA repeat number between the *URA3* insertion and the rDNA boundary in *pch2Δ* tetrads that had undergone crossover recombination. **j**, CHEF analysis of two tetrads that have undergone unequal recombination with parental controls. DNA was digested with XhoI and probed with *NTS1*. Par., parental controls; Rec., recombinants.

However, the prevalence of allelic recombinants (30%) and the fact that changes in repeat number encompassed less than 20% of the approximately 100–110 rDNA repeats in our strains indicate that the homology search for DSB repair in the outermost rDNA repeats in *pch2Δ* cells is restricted to close neighbours. Notably, the distribution of changes in repeat number closely matched the pattern of crossover events (Fig. 1h, i). This congruence indicates that the spread in the crossover distribution (Fig. 1g, h) can largely be accounted for by non-allelic exchanges between rDNA repeats originating from DSBs in the outermost repeats of the array. Finally, although rDNA exchanges occurred at a much lower frequency in wild-type cells, they were also associated with NAHR (Supplementary Table 2), indicating that Pch2 primarily acts to prevent NAHR by suppressing DSB formation. These results establish the rDNA borders as high-risk regions for meiotic NAHR.

To determine how Pch2 suppresses DSB formation near the rDNA, we measured the chromosome association of DSB-related factors at the time of DSB formation. The three essential DSB factors[11,12] that we were able to analyse by chromatin immunoprecipitation (ChIP), Rec114, Mer2 and Mre11, were specifically enriched near the rDNA and *HML* in *pch2Δ* cells, mirroring the changes in DSB formation (Fig. 2 and Supplementary Fig. 3). We then investigated whether the regional exclusion of DSB factors could be explained by local depletion of the DSB-promoting chromosome-axis protein Hop1, the cytological distribution of which is affected by Pch2 (refs 4, 13). Although Hop1 binding was slightly increased near the rDNA and *HML* in *pch2Δ* cells, it was abundant even in wild-type cells, indicating that Pch2 does not regulate the initial chromosomal recruitment of Hop1. Rather, the differences in Hop1 binding that we observed might reflect an effect of Pch2 on chromosome structure[13]. Finally, although DSBs are enriched in promoters containing histone H3 lysine 4 trimethylation (H3K4me3)[14], we saw no difference in the genome-wide levels of this modification with or without Pch2 (Fig. 2c and Supplementary Figs 3d and 4), indicating that Pch2 does not influence this chromatin modification. These findings indicate that Pch2 specifically blocks the stable recruitment of DSB factors to prevent local DSB formation.

We sought to identify proteins that collaborate with Pch2 in preventing rDNA-proximal DSBs. A yeast two-hybrid screen isolated a fragment of the Orc1 protein, containing its ATPase domain, as a Pch2 interactor (Fig. 3a). This interaction was confirmed by co-immunoprecipitation (Fig. 3b and Supplementary Fig. 5a). Orc1 is a component of the conserved origin recognition complex that has several important chromosomal functions, including the loading of the replicative helicase[15]. Impairing Orc1 protein levels by a temperature-sensitive *orc1-161* mutation[16] (Supplementary Fig. 5b) triggered DSB formation in the rDNA flanking regions, similarly to the loss of *PCH2* (Fig. 3c, d). DSB formation near the rDNA occurred even at a temperature (23 °C) that is permissive for pre-meiotic DNA replication and spore viability (Fig. 3c–e and Supplementary Fig. 5c, d). Similarly, we saw increased DSB levels near the rDNA in an *orc1* mutant lacking the N-terminal bromo-adjacent homology (BAH) domain that is required for the chromatin-silencing function of Orc1 (ref. 17), but is dispensable for DNA replication (Fig. 3f and Supplementary Fig. 5e, f). These data indicate that the regulatory roles of Orc1 in DSB formation and bulk DNA replication are separable, although we cannot rule out that the analysed *orc1* mutations affect rDNA replication locally. During meiosis, Pch2 concentrates in the nucleolus, the organelle assembled on the rDNA array[4]. In *orc1-161* cells, the recruitment of Pch2 to the nucleolus was impaired, despite normal levels of cellular Pch2 (Fig. 3g, h). Both Pch2 and Orc1 belong to the AAA+ family of ATPases that often function as multimeric complexes[18], and we found that the ATPase activity of Pch2 was required to prevent rDNA-proximal DSBs (Supplementary Fig. 5g). These data define a role for Orc1 in the nucleolar recruitment and possible activation of Pch2 to prevent local DSB formation.

To find out whether the specific DSB activity at the rDNA borders in *pch2Δ* mutants was linked to the presence of the rDNA itself, we
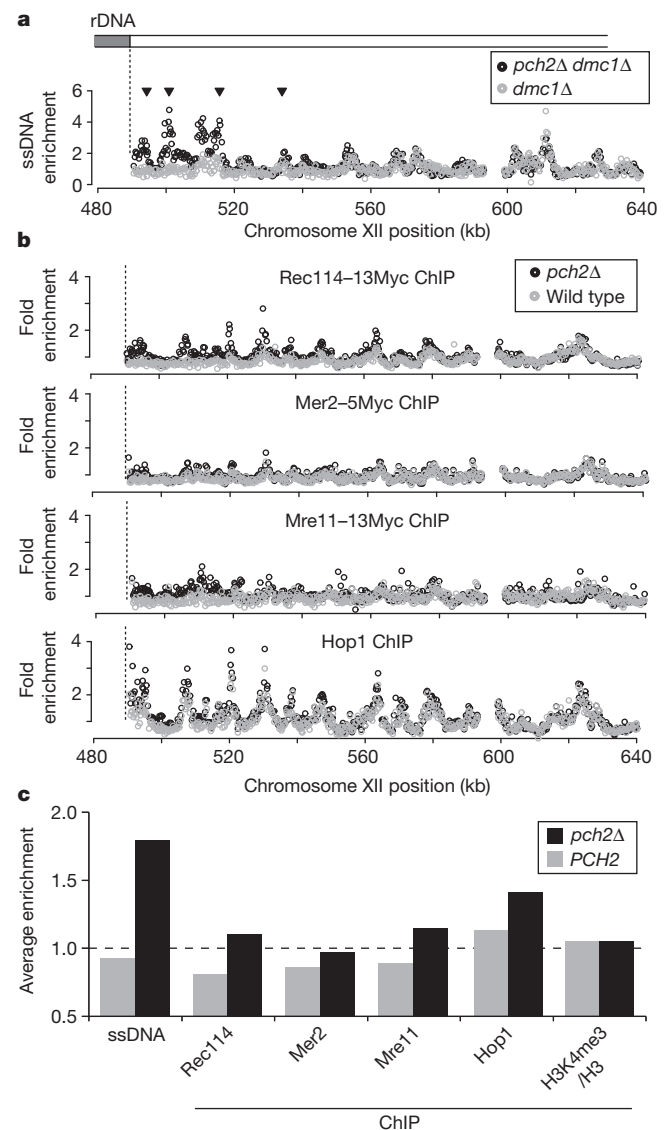
Figure 2 | Association of the meiotic DSB machinery near the rDNA.
a, Enrichment of ssDNA in a region flanking the right border of the rDNA on chromosome XII (*Saccharomyces* genome database (SGD) coordinates) in *dmc1Δ* (H118, grey) and *pch2Δ dmc1Δ* (H2629, black) cells. Arrowheads indicate >twofold increased DSB formation in *pch2Δ dmc1Δ* compared to *dmc1Δ* cells. **b**, ChIP-chip analysis in same region as in **a**, for the proteins Rec114–13Myc (first panel; wild type (H4890), *pch2Δ* (H4893)), Mer2–5Myc (second panel; wild type (H5917), *pch2Δ* (H5916)), Mre11–13Myc (third panel; wild type (H5547), *pch2Δ* (H5947)) and Hop1 (fourth panel; wild type (H119), *pch2Δ* (H2817)), in wild-type (grey) and *pch2Δ* (black) cells. **c**, Average enrichment for the different *PCH2* and *pch2Δ* ssDNA and ChIP data sets (see **a**, **b** and Supplementary Fig. 4) within the 50 kb flanking the right rDNA border (see Methods for genomic coordinates). The genome-wide average is indicated by the dotted line.

deleted the rDNA array from its genomic location. In these strains, loss of *PCH2* no longer allowed DSB formation in the flanking regions (Figs 4a, b), demonstrating an intrinsic DSB-promoting activity in the rDNA. To investigate whether the rDNA was sufficient to promote DSB formation, we created a translocation between chromosomes II and XII that exchanged the rDNA-distal portion of chromosome XII with a portion of chromosome II (Fig. 4c). In these strains, DSB levels were no longer increased in the former right flank of the rDNA (now flanked by chromosome II sequences; Fig. 4d), but notably, increased DSB formation was observed on the chromosome II sequences that, after translocation, were flanking the rDNA (Fig. 4e). Thus, in the absence of *PCH2*, the rDNA is necessary and sufficient to promote DSB formation.
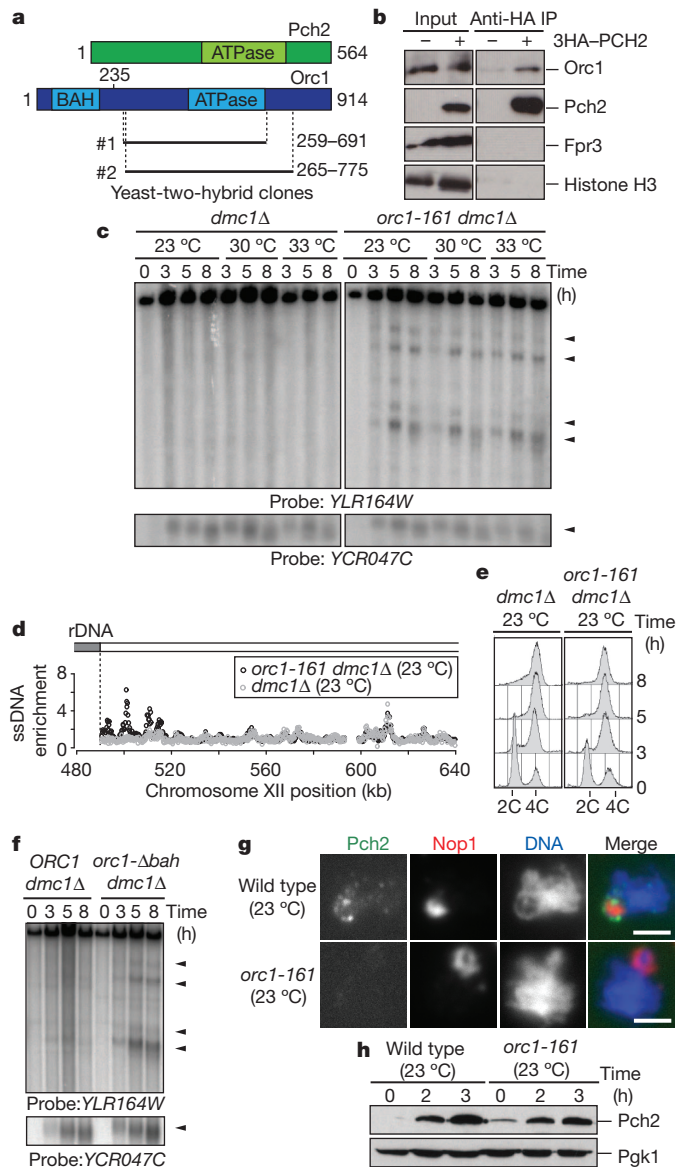
Figure 3 | Orc1 and Pch2 collaborate to suppress DSB formation.
a, Schematic of Pch2 and Orc1 proteins, indicating Orc1 clones identified by yeast two-hybrid screen. **b**, Co-immunoprecipitation (IP) between haemagglutinin-tagged Pch2 (3HA–Pch2) and Orc1 in wild-type (H119) and *3HA–PCH2* (H3463) cells. Fpr3 and histone H3 are controls for the nucleolar and chromosomal fractions, respectively. **c**, Southern blots of the right rDNA flank and *YCR047C* in meiotic *dmc1Δ* (H118) and *orc1-161 dmc1Δ* (H4952) cells, grown at the indicated temperatures. Arrowheads indicate broken DNA fragments. **d**, Profiles of ssDNA in the region flanking the right rDNA border in *orc1-161 dmc1Δ* (H5137, black) and *dmc1Δ* (H118, grey) cells, grown at 23 °C. **e**, DNA-content analysis of meiotic *dmc1Δ* (H118) and *orc1-161 dmc1Δ* (H5137) cells grown at 23 °C. 2C and 4C refer to unreplicated and replicated diploid DNA contents, respectively. **f**, Southern blots of the right rDNA flank and *YCR047C* in *ORC1 dmc1Δ* (H5838) and *orc1-Δbah dmc1Δ* (H5865) cells. **g**, Immunofluorescence of chromosome spreads stained for Pch2 (HA, green), Nop1 (nucleolar marker, red) and DNA (blue) in *3HA–PCH2* (H3463) and *3HA–PCH2 orc1-161* (H5033) cells at 3 h after meiotic induction at 23 °C. Scale bar, 2 μm. **h**, Western blot analysis showing Pch2 (HA) expression in *3HA–PCH2* (H3463) and *3HA–PCH2 orc1-161* (H5033) cells. Pgk1 is used as loading control.

The rDNA is assembled into specialized, Sir2-dependent chromatin, and we asked how this chromatin state influenced DSB formation at the rDNA boundaries. Notably, loss of Sir2 protein or loss of its deacetylase activity[19] largely eliminated DSB formation in the rDNA flanking regions in *pch2Δ* mutants (Fig. 4f, g). Thus, although
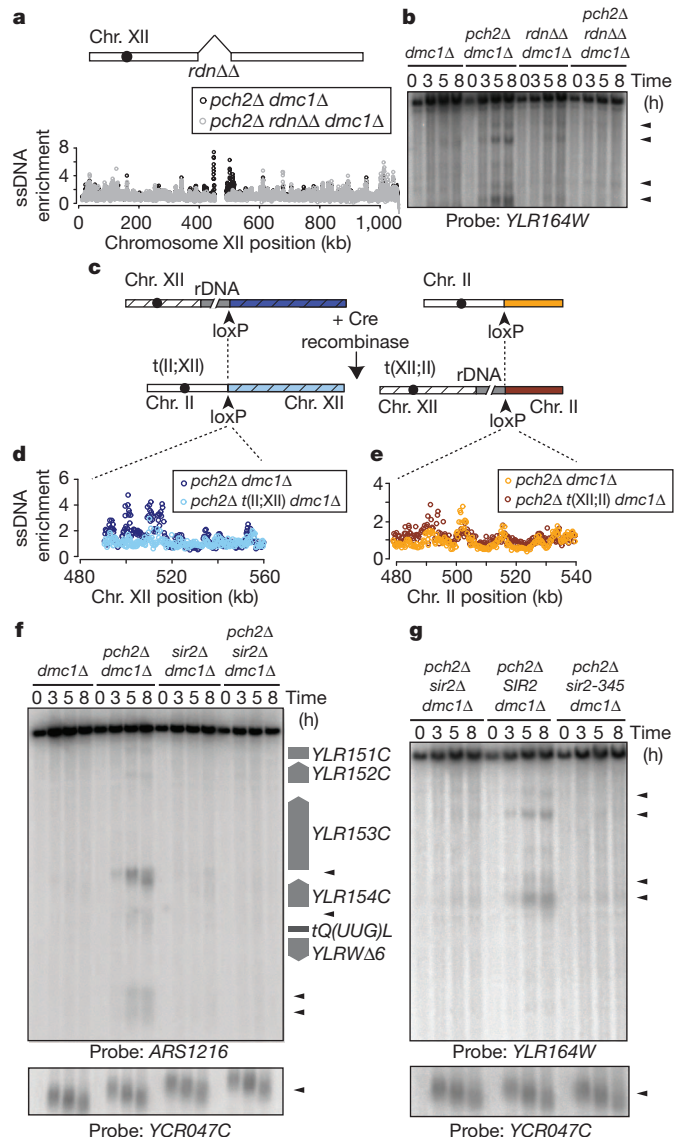
Figure 4 | Ribosomal-DNA chromatin promotes DSB formation.
a, Schematic of the rDNA deletion strain (*rdnΔΔ*) and ssDNA profiles of the region flanking the right rDNA border in *pch2Δ dmc1Δ* (H2629, black) and *pch2Δ rdnΔΔ dmc1Δ* (H4737, grey) cells. **b**, Southern blot analysis of the right rDNA flank in *dmc1Δ* (H118), *pch2Δ dmc1Δ* (H2629), *rdnΔΔ dmc1Δ* (H4736) and *pch2Δ rdnΔΔ dmc1Δ* (H4737) cells. Arrowheads indicate broken DNA fragments. **c**, Strategy used to generate chromosomal translocations between chromosomes XII and II. **d**, **e**, Profiles of ssDNA in strains containing the XII;II translocation. In **d**, the depicted region is next to the rDNA in *pch2Δ dmc1Δ* (H2629) cells (dark blue) and next to the left arm of chromosome II in *pch2Δ* t(II;XII) *dmc1Δ* (H4798) cells (light blue). In **e**, the depicted region is located on chromosome II in *pch2Δ dmc1Δ* cells (orange) and next to rDNA in *pch2Δ* t(XII;II) *dmc1Δ* cells (dark red). **f**, Southern blot of the left rDNA flank (HindIII digest; *ARS1216* probe) and of *YCR047C* in *dmc1Δ* (H118), *pch2Δ dmc1Δ* (H2629), *sir2Δ dmc1Δ* (H2953) and *pch2Δ sir2Δ dmc1Δ* (H3038) cells. Positions of open reading frames are shown schematically alongside the Southern blot. **g**, Southern blot of the right rDNA flank and *YCR047C* in *pch2Δ sir2Δ dmc1Δ* (H3262), *pch2Δ sir2Δ dmc1Δ leu2::SIR2* (H3261) and *pch2Δ sir2Δ dmc1Δ leu2::sir2-345* (H3282) cells. Arrowheads indicate broken DNA fragments.

Sir2-dependent heterochromatin suppresses meiotic DSBs within the rDNA array (refs 2, 3 and Fig. 1a), it has a profound DSB-promoting effect on the rDNA borders that is counteracted by Pch2 and Orc1. It is also notable that Sir2 itself localizes Pch2 to the nucleolus[4], reflecting an elegant coupling mechanism that maintains meiotic stability across the entire rDNA. The double dependence of Pch2 on Sir2 and Orc1

may promote Pch2 enrichment at the nucleolus, analogous to the bimodal recruitment mechanisms that restrict localization of Aurora B and shugoshin to centromeres[20].

Although repeat-associated chromatin marks differ substantially between organisms and even among individual loci[21], the assembly of heterochromatin on repetitive DNA arrays is a common strategy to protect the genome against destabilization caused by errors in meiotic recombination[1]. Our results establish borders between heterochromatin and euchromatin as potential high-risk regions for meiotic DSB formation and NAHR, and reveal the existence of a secondary border-specific system that shields against these events. Buffer zones like those established by Pch2 and Orc1 may need to be broad, because even DSBs adjacent to repetitive DNA can trigger NAHR[22]. Given the prominent presence of repetitive DNA arrays in genomes ranging from yeast to man[1], we propose that mechanisms that limit DSB activity around repetitive DNA might be a widespread phenomenon.

METHODS SUMMARY

Yeast strains were of the SK1 background and are listed in Supplementary Table 3. Analysis of single-stranded DNA profiles and ChIP-chip analysis were performed as previously described[6,23]. These and other standard techniques used are detailed in the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Sasaki, M., Lange, J. & Keeney, S. Genome destabilization by homologous recombination in the germ line. *Nature Rev. Mol. Cell Biol.* **11**, 182–195 (2010).
2. Gottlieb, S. & Esposito, R. E. A new role for a yeast transcriptional silencer gene, *SIR2*, in regulation of recombination in ribosomal DNA. *Cell* **56**, 771–776 (1989).
3. Mieczkowski, P. A. *et al.* Loss of a histone deacetylase dramatically alters the genomic distribution of Spo11p-catalyzed DNA breaks in *Saccharomyces cerevisiae. Proc. Natl Acad. Sci. USA* **104**, 3955–3960 (2007).
4. San-Segundo, P. A. & Roeder, G. S. Pch2 links chromatin silencing to meiotic checkpoint control. *Cell* **97**, 313–324 (1999).
5. Wu, H. Y. & Burgess, S. M. Two distinct surveillance mechanisms monitor meiotic chromosome metabolism in budding yeast. *Curr. Biol.* **16**, 2473–2479 (2006).
6. Blitzblau, H. G. *et al.* Mapping of meiotic single-stranded DNA reveals double-stranded-break hotspots near centromeres and telomeres. *Curr. Biol.* **17**, 2003–2012 (2007).
7. Gerton, J. L. *et al.* Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae. Proc. Natl Acad. Sci. USA* **97**, 11383–11390 (2000).
8. Petes, T. D. Meiotic recombination hot spots and cold spots. *Nature Rev. Genet.* **2**, 360–369 (2001).
9. Keeney, S. Mechanism and control of meiotic recombination initiation. *Curr. Top. Dev. Biol.* **52**, 1–53 (2001).
10. Coelho, P. S. *et al.* A novel mitochondrial protein, Tar1p, is encoded on the antisense strand of the nuclear *25S* rDNA. *Genes Dev.* **16**, 2755–2760 (2002).
11. Arora, C., Kee, K., Maleki, S. & Keeney, S. Antiviral protein Ski8 is a direct partner of Spo11 in meiotic DNA break formation, independent of its cytoplasmic role in RNA metabolism. *Mol. Cell* **13**, 549–559 (2004).
12. Keeney, S. & Neale, M. J. Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. *Biochem. Soc. Trans.* **34**, 523–525 (2006).
13. Borner, G. V., Barot, A. & Kleckner, N. Yeast Pch2 promotes domainal axis organization, timely recombination progression, and arrest of defective recombinosomes during meiosis. *Proc. Natl Acad. Sci. USA* **105**, 3327–3332 (2008).
14. Borde, V. *et al.* Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *EMBO J.* **28**, 99–111 (2009).
15. Bell, S. P. The origin recognition complex: from simple origins to complex functions. *Genes Dev.* **16**, 659–672 (2002).
16. Gibson, D. G., Bell, S. P. & Aparicio, O. M. Cell cycle execution point analysis of ORC function and characterization of the checkpoint response to ORC inactivation in *Saccharomyces cerevisiae. Genes Cells* **11**, 557–573 (2006).
17. Bell, S. P. *et al.* The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. *Cell* **83**, 563–568 (1995).
18. Hanson, P. I. & Whiteheart, S. W. AAA+ proteins: have engine, will work. *Nature Rev. Mol. Cell Biol.* **6**, 519–529 (2005).
19. Imai, S., Armstrong, C. M., Kaeberlein, M. & Guarente, L. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795–800 (2000).
20. Vader, G. & Lens, S. M. Chromosome segregation: taking the passenger seat. *Curr. Biol.* **20**, R879–R881 (2010).
21. Moazed, D. Common themes in mechanisms of gene silencing. *Mol. Cell* **8**, 489–498 (2001).
22. Hoang, M. L. *et al.* Competitive repair by naturally dispersed repetitive DNA during non-allelic homologous recombination. *PLoS Genet.* **6**, e1001228 (2010).
23. Aparicio, O. M., Weinstein, D. M. & Bell, S. P. Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. *Cell* **91**, 59–69 (1997).

METHODS

Yeast strains and two-hybrid analysis. All yeast strains used in this study were constructed in the SK1 background and are listed in Supplementary Table 3. Epitope tags and gene disruptions were introduced by standard PCR-based transformation. The *orc1-Δbah* mutant was generated using a plasmid encoding a truncated version of Orc1 (amino acids 235–914; pSPB1.48, gift from S. P. Bell[24]). To create *URA3* insertions in the rDNA, cells were transformed with a pRS306-NTS1/2 plasmid linearized with SphI (this plasmid contains a 2341-bp fragment harbouring the intergenic rDNA sequences, *NTS1* and *NTS2*, ligated into the BamHI and EcoRI sites of pRS306). Insertion sites in the rDNA were mapped by CHEF gel analysis using a unique XhoI site in the inserted sequence (XhoI does not cut in the rDNA), and suitable clones were selected for further analysis. SK1 strains lacking the rDNA array were generated as described in ref. 25. Briefly, cells were transformed with a very high-copy rDNA plasmid (*pRDN-hyg::URA3::leu2-8*) carrying a recessive point mutation that confers resistance to hygromycin[26]. After selection on hygromycin, a clone was selected that had lost all but three repeats of the rDNA array through spontaneous deletion, as determined by CHEF gel analysis. The remaining rDNA copies were subsequently deleted by conventional gene disruption using a *HIS3* deletion cassette. Complete deletion of the rDNA array was confirmed by Southern blotting. Chromosomal translocations between chromosomes XII and II were generated essentially as previously described[27]. Briefly, plasmids containing a promoter-less *ADE2* gene adjacent to a *loxP* site (*loxP-ADE2::natMX4*) and a *GPD* promoter (*pGPD*) with an adjacent *loxP* site (*pGPD-loxP::hphMX4*; both plasmids were gifts from N. Hunter) were integrated at *YLR162W-A* and *LYS2*, respectively. After induction of Cre recombinase from a pGAL-Cre plasmid (N. Hunter), cells were selected that had undergone translocation between *LYS2* (chromosome II) and *YLR162W-A* (chromosome XII). Translocation was confirmed by Southern blot analysis. To identify interactors of Pch2 by two-hybrid screen, full-length *PCH2* was amplified from genomic DNA and the intron was removed by site-directed mutagenesis. The Pch2 coding sequence was then cloned into pGBDU-C1, and the resulting bait plasmid was used to screen libraries in all three reading frames[28].

Synchronous meiosis. Cells were grown for 24 h in yeast peptone dextrose (YPD) at 23 °C then diluted in BYTA medium (1% yeast extract, 2% tryptone, 1% potassium acetate, 50 mM potassium phthalate) to a optical density at 600 nm (OD$_{600}$) of 0.3 (OD$_{600}$ = 0.5 for *orc1-161*, *rdnΔΔ* and XII;II translocation strains), and grown for 16 h at 30 °C (or for 18 h at 23 °C in the case of temperature-sensitive strains). After two washes in water, cells were diluted into SPO medium (0.3% potassium acetate) at OD$_{600}$ = 1.9 and incubated at 30 °C unless otherwise stated.

Isolation of ssDNA. For ssDNA analysis[29,30], about 10^9 cells were fixed in 70% ethanol at −20 °C at 0 h and 5 h after induction of meiosis. After spheroplasting in sorbitol buffer (1 M sorbitol, 1% β-mercaptoethanol, 0.2 mg ml^{-1} zymolyase, and 0.1 M EDTA, pH 7.4), cells were lysed in NDS buffer (0.6% SDS, 300 mM EDTA, 10 mM Tris-HCl, pH 9.5). After treatment with proteinase K (0.25 mg ml^{-1}) and RNase A, the DNA was digested with EcoRI and ssDNA was then enriched by adsorption to BND-cellulose and eluted using 1.8% caffeine. This enriched ssDNA was subsequently used for microarray analysis. For this, 1.5 μg of the respective 0-h and 5-h ssDNA samples was labelled with Cy3-dUTP or Cy5-dUTP (GE Healthcare) by random priming without denaturation using 4 μg of random primer (Integrated DNA Technologies) and 10 units of Klenow (New England Biolabs).

Western blotting and immunoprecipitation. For western blotting, 5 ml of meiotic cells were harvested at the indicated time points and resuspended in 5% trichloroacetic acid. After incubation on ice for 10 min, samples were washed in acetone and dried overnight. Samples were lysed by bead beating in a FastPrep FP120 (Thermo Scientific) in TE lysis buffer (10 mM Tris (pH 7.5), 1 mM EDTA, 2.75 mM dithiothreitol). SDS loading buffer (3×) was added and the pH of the sample was adjusted to neutral by addition of 1 M Tris (pH 8.0). Samples were separated by standard polyacrylamide gel electrophoresis. The following antibodies were used: anti-ORC (1108, 1:1,000, gift from S. P. Bell), anti-Fpr3 (1:1,000, gift from J. Thorner), anti-HA (3F10, 1:1,000, Roche), anti-Pgk1 (1:1,000, Invitrogen) and anti-histone H3 (1:1,000, Abcam). For immunoprecipitations, 50 ml of meiotic cells were harvested 3 h after induction of meiosis. Cells were diluted in 2× lysis buffer (20 mM HEPES (pH 7.5), 4 mM MgCl$_2$, 0.6 M glutamic acid, 0.32 M sorbitol, 4% glycerol, 0.5% Triton X-100) containing protease inhibitors, and lysed by bead beating. Extracts were sonicated and cleared by centrifugation. After removal of one tenth of the extract for an input sample, extracts were immunoprecipitated with 2 μl anti-HA (3F10, Roche) for 3HA–Pch2, 2 μl anti-ORC (1108) and 2 μl anti-Fpr3, in combination with 20 μl of a 50% slurry of GammaBind-Sepharose beads (GE Healthcare) for 16 h at 4 °C. After five washes in 1× lysis buffer, 1× SDS loading buffer was added to the beads, and samples were analysed by western blotting with the indicated antibodies.

Chromatin immunoprecipitation. Meiotic cells (25 ml) were harvested 3 h after induction of meiosis and fixed for 15 min in 1% formaldehyde. The formaldehyde was quenched by addition of 125 mM glycine. Samples were processed as previously described[23]. Before immunoprecipitation, one tenth of the sample was removed as input sample. The antibodies used for immunoprecipitation were: 2 μl anti-Myc (9E11, Abcam; for Rec114–13Myc, Mer2–5Myc and Mre11–13Myc), 2 μl anti-Hop1 (gift from N. Hollingsworth), 2 μl anti-histone-H3 (AB1791, Abcam) and 2 μl anti-H3K4me3 (AB8580, Abcam), in combination with 20 μl of a 50% slurry of GammaBind-Sepharose beads (GE Healthcare). Half of the ChIP sample and one tenth of the input were labelled with Cy3-dUTP or Cy5-dUTP (GE Healthcare) as described in the ssDNA protocol, with the difference that the DNA was denatured for 5 min at 95 °C before the extension reaction.

Microarray analysis. After removal of unincorporated dyes, Cy3- and Cy5-labelled samples were hybridized to custom 4 × 44K tiled genomic yeast microarrays (Agilent Technologies) for 16 h at 65 °C. Levels of Cy3 and Cy5 were calculated with the Agilent Feature Extractor CGH software. Background normalization, log$_2$ ratios for each experiment and scale normalizations between experiments were calculated with the sma package in R (v2.1.0, http://www.r-project.org)[29,30]. Each data set is an average of two experiments. For comparison between isogenic wild-type and *pch2Δ* cells, data sets were scale-normalized. To analyse the distribution of H3K4me3, we normalized the enrichment to that of total histone H3 generated from the same extracts, by subtracting the log$_2$ ratios. To measure the average ssDNA or ChIP enrichment in different chromosomal regions, the following SGD coordinates were analysed, on the basis of the positions of available array features:

50 kb right of the rDNA: XII, 490,531–540,530.
50 kb left of the rDNA: XII, 401,371–451,370.
Rest of chromosome XII: XII, 1–401,370 and 540,531–1,078,177.
First 100 kb of chromosome III: III, 1–100,000.
Rest of chromosome III: III, 100,001–316,620.
Chromosome VIII: VIII, 1–562,643.

Chromosome spreads and immunofluorescence. Meiotic cells were spread as described previously[31]. Cells were spheroplasted at 37 °C in solution 1 (2% potassium acetate, 0.8% sorbitol, 10 mM dithiothreitol, 130 mg ml^{-1} zymolyase 100T (Seikagaku)). Solution 2 (100 mM MES (pH 6.4), 1 mM EDTA, 0.5 mM MgCl$_2$, 1 M sorbitol) was added to stop spheroplasting. Spheroplasted cells (15 μl) were fixed with 30 μl of fixative solution (4% paraformaldehyde, 3.4% sucrose) and lysed with 60 μl 1% lipsol. After addition of 60 μl fixative solution, cells were spread using a glass rod. After drying, the slides were blocked in blocking buffer (0.2% gelatine, 0.5% BSA in PBS) and stained with the following antibodies: anti-HA (3F10, 1:500 dilution, Roche), and anti-Nop1 (1:500 dilution, Encor Biotechnology).

CHEF gel electrophoresis and Southern blotting. Chromosome fragments for CHEF analysis were prepared by restriction digest in agarose plugs. Briefly, 20 ml of meiotic cells were killed by addition of sodium azide (0.1% final w/v), pelleted and stored on ice for the duration of the time course. Cell pellets were washed twice in CHEF-TE (10 mM Tris-HCl (pH 7.5), 50 mM EDTA) and resuspended in 300 μl CHEF-TE. Tubes were individually treated as follows: 4 μl zymolyase T100 (10 mg ml^{-1}) was added and the mix was incubated at 42 °C for 30 s before addition of 500 μl low-melting-point agarose (1.5% SeaPlaque GTG, 125 mM EDTA) at 42 °C. Gel plugs (90 μl) were allowed to harden on ice in disposable plug molds (Bio-Rad) and incubated overnight at 37 °C in 300 μl LET (10 mM Tris (pH 7.5), 500 mM EDTA) per plug. Plugs were deproteinized overnight at 50 °C in 200 μl NDS-PK (LET, 1% *N*-lauroylsarcosine, 1 mg ml^{-1} proteinase K (Amresco)) per plug. Proteinase K was inactivated by incubating plugs for 1 h at 4 °C in CHEF-TE containing 1 mM PMSF, and plugs were washed three times in CHEF-TE, then digested with XhoI in digestion buffer containing 5 mM spermidine. To analyse the entire rDNA array, digested chromosomes were separated by CHEF gel electrophoresis in 1% agarose in 0.5× TBE, 6 V cm^{-1}, using 60-s pulses for 15 h and 90-s pulses for 9 h. For fine mapping of rDNA insertions and to analyse changes in repeat number, digested chromosome fragments were separated using a 5–20 s ramp over 20 h. For conventional electrophoresis, DNA fragments were separated on 0.6% agarose in 1× TBE and transferred onto Hybond-XL membranes (GE Healthcare) by alkaline transfer. Southern blotting was performed as previously described[32] and quantified with a Fujifilm BAS-2500 image reader V1.8 and Multi Gauge V2.2 software.

Probes for Southern analysis. Probe templates for non-rDNA sequences were generated by nested PCR and gel purification. The following probes (SGD coordinates) were used:

YLR164W: XII, 493,432–493,932.
YCR047C: III, 209,361–210,030.
ARS1216: XII, 450,407–451,150.
YLR152C: XII, 443,849–444,910.
NTS1: XhoI and XbaI digest of pRS306-NTS1/2. This probe detects all *NTS1* and *NTS2* sequences (pan-rDNA probe).

rDNA insertion: BciVI digest of pRS306-NTS1/2. This probe specifically detects the plasmid backbone of the *URA3* insertion cassette.

TOM1/YDR457W: IV: 1,370,714–1,371,733

Flow cytometry. At the indicated time points, 150 µl of meiotic cells were fixed for 2 h at 4 °C after addition of 350 µl absolute ethanol. Cells were resuspended in 500 µl of 50 mM sodium citrate containing 0.7 µl RNase A (30 mg ml^{-1}, Sigma). Cells were incubated for 2 h at 50 °C. 10 µl proteinase K (Amresco) was added and cells were deproteinated for 2 h at 50 °C. 500 µl of 50 mM sodium citrate containing 0.2 µl Sytox Green (Amersham) was added to the cells. Cells were briefly sonicated and analysed using a FACScalibur (Becton Dickinson) flow cytometer. DNA profiles were generated using CellQuest software.

Recombination mapping. To determine crossover recombination rates, cells were sporulated for 24 h in 3 ml SPO and treated with zymolyase (1 mg ml^{-1} in 1 M sorbitol) to remove ascus walls. Tetrads were dissected by micromanipulation and marker segregation was determined by replica plating on appropriate selective media. For mapping within the rDNA, only tetratypes were used to calculate recombination rates, to avoid distortions originating from non-parental ditypes that were probably the result of previous mitotic recombination.

24. Bell, S. P. *et al.* The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. *Cell* **83,** 563–568 (1995).

25. Kobayashi, T. Strategies to maintain the stability of the ribosomal RNA gene repeats—collaboration of recombination, cohesion, and condensation. *Genes Genet. Syst.* **81,** 155–161 (2006).

26. Chernoff, Y. O., Vincent, A. & Liebman, S. W. Mutations in eukaryotic 18S ribosomal RNA affect translational fidelity and resistance to aminoglycoside antibiotics. *EMBO J.* **13,** 906–913 (1994).

27. Peoples, T. L. *et al.* Close, stable homolog juxtaposition during meiosis in budding yeast is dependent on meiotic recombination, occurs independently of synapsis, and is distinct from DSB-independent pairing contacts. *Genes Dev.* **16,** 1682–1695 (2002).

28. James, P., Halladay, J. & Craig, E. A. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* **144,** 1425–1436 (1996).

29. Blitzblau, H. G. *et al.* Mapping of meiotic single-stranded DNA reveals double-stranded-break hotspots near centromeres and telomeres. *Curr. Biol.* **17,** 2003 (2007).

30. Blitzblau, H. G. & Hochwagen, A. Genome-wide detection of meiotic DNA double-strand break hotspots using single-stranded DNA. *Methods Mol. Biol.* **745,** 47–63 (2011).

31. Loidl, J., Nairz, K. & Klein, F. Meiotic chromosome synapsis in a haploid yeast. *Chromosoma* **100,** 221–228 (1991).

32. Hunter, N. & Kleckner, N. The single-end invasion: an asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination. *Cell* **106,** 59–70 (2001).

CORRECTIONS & AMENDMENTS

Post–traumatic stress disorder is associated with PACAP and the PAC1 receptor

Kerry J. Ressler, Kristina B. Mercer, Bekh Bradley, Tanja Jovanovic, Amy Mahan, Kimberly Kerley, Seth D. Norrholm, Varun Kilaru, Alicia K. Smith, Amanda J. Myers, Manuel Ramirez, Anzhelika Engel, Sayamwong E. Hammack, Donna Toufexis, Karen M. Braas, Elisabeth B. Binder & Victor May

Half of the data points were inadvertently omitted from the published version of Fig. 4a; the statistical analyses in the text and figure legend, however, do refer to the complete data set. The corrected figure is shown here and has been corrected in the online versions of the paper.

In addition, we present additional information to clarify two results reported in the Article regarding plasma pituitary adenylate cyclase-activating polypeptide (PACAP) levels and post-traumatic stress disorder (PTSD) symptom associations. In the Article, we reported replication of the association between PACAP levels and hyperarousal subscale, because this was the most robust association in the initial cohort. We now present the analyses separately for initial, replication and combined cohorts in Table 1. All associations but one are significant in the replication cohort. The second issue concerns potential medical confounds that could underlie the reported association. Although we do not have medical chart data on most patients, we do have responses from a health questionnaire administered during collection of trauma history and other data. We have now reanalysed the associations for the PTSD symptom scale (PSS) hyperarousal and total symptoms using subjective reports of health condition

4a
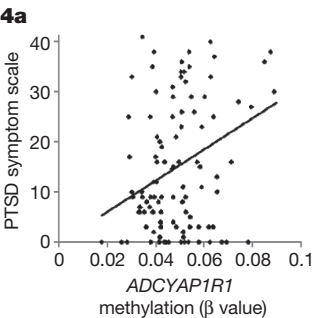
from the questionnaires as covariates. These data are presented in Table 2 and do not show any effect of health- and illness-related questions on the relationship between PACAP and PTSD symptoms. None of these additions affect the results and conclusions of the original Article.

1. *Diagnostic and Statistical Manual of Mental Disorders* 4th edn (text rev.) (American Psychiatric Association, 2000).

Table 1 | PACAP associations with PTSD symptoms

	Initial $n = 34$	Replication $n = 74$	Combined $n = 108$
**Correlation of PACAP level with PSS hyperarousal score	0.006	0.015	0.001
*High PACAP level with PSS hyperarousal score (Fig. 1c)	0.001	0.01	0.0004
*High PACAP level with PSS hyperarousal score (Fig. 1d, adjusted)	0.01	0.001	0.00005
*High PACAP level with PSS total score (Fig. 1b)	0.0003	0.15	0.008
*High PACAP level with PSS total score (adjusted)	0.004	0.04	0.002
*High PACAP level with clinically significant PTSD symptoms (Fig. 1d, adjusted)	0.0002	0.01	0.0003
*High PACAP levels with PSS-based PTSD diagnosis[1] (adjusted)	0.01	0.02	0.0008

The table shows *P* values of correlations and regression analyses for the initial, replication and combined cohort analyses: 2-tailed for initial and combined; 1-tailed for replication. 'Adjusted' means adjusted for age, substance use and trauma.
** Bivariate correlation.
* Analysis of variance (ANOVA).

Table 2 | Health-adjusted PACAP associations

	Initial $n = 28$	Replication $n = 58$	Combined $n = 86$
Health adjusted-association of high PACAP levels with PSS hyperarousal score	0.031	0.005	0.001
Health adjusted-association of high PACAP levels with PSS-based PTSD diagnosis[1]	0.05	0.03	0.002

The table shows *P* values of correlations and regression for the initial, replication and combined cohort analyses: 2-tailed analyses for initial and combined; 1-tailed for replication.

CAREERS

The opening of the Scripps Research Institute's campus in Jupiter, Florida, was the start of an influx of bioscience to the state — a flow that has now slowed.

FLORIDA

Second thoughts

The sunshine state's rush to become a bioscience player started with a bang. Now it faces more realistic expectations.

BY SARAH KELLOGG

Eight years ago, Florida's officials wanted to transform the state into a bioscience powerhouse, hoping that it would become known for more than just beach-front retirement apartments and cartoon-themed amusement parks. So they launched a bold initiative to swiftly build world-class bioscience clusters in the region. The plan has cost the state and local governments more than US\$1.5 billion, sparked significant growth in Florida's once-sleepy biotechnology industry — and left many expecting more. The promise of a robust industry with new jobs, research funding, patentable discoveries and royalties has yet to be fully realized. That is frustrating for a state grappling with a lacklustre economy and severe budget shortfalls.

Nevertheless, Florida's bioscience investment has brought in some recruitment, federal funding, public–private partnerships and start-ups, albeit more slowly than many would like. "Florida was not much on the radar of the international scientific community when all this started," says Claudia Hillinger, vice-president for institute development at the Max Planck Florida Institute in Jupiter, a brain-research centre and the first US campus of the German Max Planck Society. "That has changed in the last three years. Things are starting to develop and evolve, but I think we're going to need a bit more patience from everyone," she adds.

AN AMBITIOUS START

When Florida launched its effort to promote bioscience in 2003, the US economy was thriving. Then-Governor Jeb Bush (Republican) was on a personal mission to restructure Florida's financial landscape for the twenty-first century, and the state was flush with private and public funds.

The initiative — comprising generous government subsidies and aggressive recruitment — proceeded quickly. Life-science clusters, mostly in central and southern Florida, emerged in cities commonly associated with orange groves and sandy beaches (see *Nature* **446,** 1112–1113; 2007). Respected senior researchers and promising postdocs headed to the region, their ambitions limited only by their imaginations. State and local incentives, including direct grants and tax breaks, drew prestigious research institutions. The Scripps Research Institute, a biomedical-research centre headquartered in La Jolla, California, received \$579 million in combined subsidies to open Scripps Florida in Jupiter; the Sanford-Burnham Institute for Medical Research, also headquartered in La Jolla, drew \$310 million to open a centre in Orlando; and the Max Planck Society got \$188 million for its Florida campus. Almost overnight, Florida shot from its bioscience infancy to adolescence.

The *Battelle/BIO State Bioscience Initiatives 2010* report from Battelle, an independent research firm based in Columbus, Ohio, showcases Florida's good early track record. ▶

The state ranks 6th in the country for the number of bioscience jobs held in 2008, with 27,960; and 17th in terms of US National Institutes of Health grants received in 2009, with $466 million. Florida's biotech jobs pay well, with an average salary of $55,264 in 2008, compared with $39,596 for private-sector jobs overall in the state, the report found.

Scripps Florida was the first organization to come on board, with temporary labs opening in 2005 and a dedicated campus launching in 2009. It showed that Florida was serious about bioscience, convinced other institutes to follow its example, and spurred interest in innovative partnerships with private companies hoping to exploit translational research. It currently has 40 principal investigators and 450 staff members, and plans to increase those numbers to 60 faculty members and 545 staff by 2014. It is seeking researchers with expertise in cancer biology, metabolism and ageing, and molecular therapeutics, among others. "Since we've been here, $200 million of federal money has come to us and the state," says Harry Orf, vice-president of scientific operations for Scripps Florida. "This is entirely new money coming into the state, and we continue to secure more."

HOBBLED BY THE ECONOMY

Sustaining such intense growth is difficult. Many of the public officials who championed biotech investment have left office; federal research funding has diminished; and Florida's government is strapped for cash. The state is keeping its early financial commitments to institutes, but other investments have slowed.

In 2010, the Florida legislature's Office of Program Policy Analysis and Government Accountability released a report — *Biotechnology Clusters Developing Slowly* — examining the sector. It noted that in 2008, 57% of Florida's biotech employment was in the counties hosting the research institutes, which have promised the state at least 1,100 highly paid research positions. But the report pointed out that the investment "has not yet resulted in the growth of technology clusters in the counties where program grantees have established facilities".

These conclusions reflect the ravages of high expectations and a floundering economy. "I think we're looking at a different strategy going forward," says Russell Allen, president and chief executive of BioFlorida, the state's bioscience trade association. "I don't anticipate we'll see hundreds of millions of incentive dollars any time soon. I'm not sure that would be our next best step anyway. We've proven we can build a research base. Now it's about creating jobs."

In May, Florida lawmakers approved a 2011–12 budget that allocated $3.48 billion to the state's 11 public universities — a 4% drop from the 2010–11 budget. This is the fourth year in a row that universities have seen a cut in state funding, and they have had to implement hiring freezes, tuition-fee increases and budget cuts. This year, to further trim state spending, Republican Governor Rick Scott ordered cuts of a total of $615 million from the overall proposed budget, including $6 million to build a University of Florida research facility in Lake Nona, $6 million to complete the applied science building at Florida State University in Tallahassee, and $6.3 million to build the University of Central Florida's interdisciplinary research and incubator facility in Orlando. He also vetoed $2 million in new money for obesity and diabetes research at the Sanford-Burnham Institute.

"From a funding perspective, we all wish things were different," says Ryan West, director of talent and economic development at the Florida Chamber of Commerce. "I heard no legislator take any pride or joy in reducing funding levels for any of these institutions or reducing the investment overall in the biosciences. Nobody is happy about it."

Funding anxieties seem to have contributed to the collapse of a deal in June, when, owing to a lack of state investment, the Jackson Laboratory genetic-research firm in Bar Harbor, Maine, withdrew a bid to build a lab in Sarasota County. "We were invited to submit a much-reduced proposal to the [state], but the amount available … and the uncertainty of future funding, made such a venture too speculative to undertake," says Charles Hewett, Jackson Lab's executive vice-president, in a written statement.

"Things are starting to develop and evolve, but I think we are going to need a bit more patience."
Claudia Hillinger

To encourage the private sector, the state is looking beyond direct financial support to regulatory changes and tax incentives that could foster biotech growth. In early 2011, lawmakers approved an annual corporate income-tax credit, similar to an extant federal programme, for companies that invest in research and development in Florida. "Growing this sector occurs from a lot more than providing financial support," says Stuart Doyle, a spokesman for Enterprise Florida in Orlando, the state's main economic-development agency.

A DEARTH OF VENTURE CAPITAL

The biggest potential obstacle to sustained long-term growth in Florida biotechnology is a paucity of in-state venture-capital firms willing to invest in biotech start-ups. "The state works very hard to help companies like ours, but you're just not going to find the savvy level of investor you find in California or New York," says Marilyn Bruno, founder and chief executive of Aequor, a start-up biotech company in Coral Gables. Bruno is seeking investors to develop environmentally friendly compounds that halt bacterial contamination on a wide variety of surfaces, ranging from boat hulls to skin. "Why should they invest in a risky biotech start-up they don't understand if they can buy condos on Miami Beach and make a mint?" she asks.

Florida's State Board of Administration has been authorized to invest up to 1.5% of the net assets of the state retirement-system trust fund in technology and high-growth investments. And in 2007, the state set aside $29.5 million for the Florida Opportunity Fund, created to underwrite in-state venture-capital firms. It has committed most of its resources to companies and hedge funds that support information technology, homeland security, defence and biotech start-ups in the state. But allowing the money to be used so broadly has diluted the fund's influence, says Bruno.

PORTENDING PROMISE

Despite the obstacles, Florida's bioscience efforts are bearing fruit. A cluster is growing in Orlando, where the Sanford-Burnham Institute opened an $85-million building in May; it anchors the Lake Nona Medical City, including the University of Central Florida College of Medicine, Nemours Children's Hospital and the Veterans' Administration Medical Center, both slated to open in 2012. Once completed, the Medical City is expected to employ some 30,000 scientists, administrative staff and others.

"If you simply look at Sanford-Burnham employees, about 300 people, that doesn't seem like many jobs," says Daniel Kelly, scientific director for the institute. But "the point is we're part of the engine that runs this entire cluster. It's important for credibility," he says. "It's important for investment."

Cooperation between Florida universities and institutes has helped to draw recruits from more established entities, says David Fitzpatrick, chief executive and scientific director of Max Planck Florida, who left Duke University in Durham, North Carolina, for his Florida post in January this year. "It wasn't easy for me, leaving Duke," says Fitzpatrick. "But this was an opportunity to set up an institute and to take it in a direction that I think is very exciting." Max Planck Florida expects to employ a total of 150 staff members by 2015, with 12 directors and research-group leaders, 114 other scientific staff members and 24 in administration.

Ultimately, the newness of Florida's biosciences sector is both its greatest selling point and a notable hurdle, say observers. Regardless of early and continuing growth, Florida's bioscience clusters are untested commodities, and it will take years to determine whether the state has turned a strategic investment into a flourishing industry. ∎

Sarah Kellogg *is a freelance writer in Washington DC.*

MAX PLANCK FLORIDA INST.

COLUMN

The magical world of data sharing

Andrew Peterman says scientists need to reach out.

Backstage at Disneyland, the scene is not pretty. Think of Bambi's mother dying or Simba's father being killed in the wildebeest stampede — those heart-wrenching moments when the magic of your childhood starts to waver. Even though I was well into adulthood, that's how I felt the first time I did backstage fieldwork at Disneyland. The enchantment slipped away as I watched the characters of my childhood tear their stuffed heads off, dripping in sweat.

I recalled this experience three years ago, as I sat in my office at Walt Disney Imagineering — the science and technology division of the Walt Disney Company, based in Glendale, California. The song *It's a Small World* played as I waited on hold for a colleague to check for energy-use data that I had requested. These data could help our scientific team to understand how people use buildings, and to make designs more efficient. We knew that closing the exterior doors of a retail building would reduce energy consumption, but we did not know by how much or how it would, say, affect traffic flow. My role as a scientist was to research and devise strategies to reduce energy consumption across Disney.

But I and my team fought unsuccessfully to get company employees to divulge energy-use data. Most researchers have dealt with this problem, whatever their scientific field. Why, I wondered, must accessing data be such a struggle? I also wondered why Disney could not find better hold music.

My colleague, a manager at one of Disney's retail locations, came back on the line. She would not release the data, and gave no explanation. I explained that the research would remain in the company. I urged her to reconsider, given that the information could ultimately save the company money. "Sorry, no can do," she said. "Have a magical day."

It is not surprising that most people are afraid to relinquish data, even internally. Releasing data is like letting guests see Mickey Mouse tear his own head off backstage. If people see how things work, they might not want to come back. The holders of the data might also worry that exposing the information will get them into trouble. This fear is a major challenge for scientists attempting to do research that might very well help the data holders.

After more than a year of trying to gather data at Disney, despite cajoling, harassing and coming as close as I could to actually bribing facility managers, I managed it for only a few buildings. Initially, I saw this as an organizational failure for the company. But I realized that the problem was not Disney's alone. As scientists and engineers, we often assume that our research goals are the same as the goals of the people from whom we need to get the data, when in fact they might be quite different. We are not trained to effectively engage others in our research.

In pursuing my PhD, I have learned that my job is not just to research, collect and analyse data and present results. Scientists must be intimately involved in working with those who possess and control data, beyond simply extracting information. We must teach those who will be most affected by our work how and why they should be involved in the research process. And it is crucial that we explore their goals, try to understand their apprehension and work to allay those fears.

I should have tried to understand how my research affected that Disney manager — perhaps buildings with proper energy-use practices would reduce costs or improve her employees' comfort. Scientists and engineers often encounter resistance from the people who stand to benefit most from our work. It should not be that way. As the song goes, "There's so much that we share, that it's time we're aware, it's a small world after all." ∎

Andrew Peterman is a doctoral candidate in civil and environmental engineering at Stanford University in California.

Teaching aids research

Teaching others helps science graduate students to improve their own research skills, according to a study (D. F. Feldon *et al. Science* **333,** 1037–1039; 2011). The work compared science, technology, engineering or maths (STEM) graduate students who teach with those who only conduct research. It examined the quality and testability of hypotheses by the students at the start and end of an academic year, as well as the strength and design of their experiments, on the basis of assessments by independent scientific reviewers. The analysis is the first of its kind to measure the growth of skills, says lead author David Feldon, who studies STEM education at the University of Virginia in Charlottesville. He theorizes that teaching in STEM enhances early-career scientists' understanding of what comprises good research.

Degree brings prospects

About 52% of people who graduated from US professional science master's (PSM) programmes in 2010–11 had new jobs 1–6 months after earning their degrees, says a survey. *Outcomes for PSM Alumni: 2010/11*, released on 23 August by the US Council of Graduate Schools (CGS) in Washington DC, found that 39% of those with new jobs had secured them through internships associated with their PSM. CGS president Debra Stewart found the numbers encouraging. "Employers are seeing the value of the PSM," she says. Most of the jobs were research related. The survey had 320 responses from graduates of 58 PSM programmes. Advocates call the PSM a viable alternative to the PhD.

Romance beats science

Young women who want romance show less interest in science, technology, engineering and maths (STEM) than in other fields, says a study (L. E. Park *et al. Pers. Soc. Psychol. B.* **37,** 1259–1273; 2011). The authors gauged reactions of 350 students to 'romantic' images such as candlelight and sunsets, to other images of books or libraries, and to chats about dating or tests. Those who saw and heard romantic content reported less interest in STEM. Such dynamics could contribute to women's low representation in STEM, says lead author Lora Park, a psychologist at the University at Buffalo in New York.

A SENTENCE TO LIFE

Pet project.

BY IGOR TEPER

Better to be dead. Sitting on a park bench, watching the swans scribe their inscrutable trajectories on the lake's surface, streams of joggers, dog walkers, families with strollers flowing around him, but never too close, Julius felt more alone than ever. Better to be dead.

He knew he couldn't blame Tabytha, could only blame himself, but the whole thing was just so outrageous. If she hadn't broken his heart, then stomped on it; hadn't made him feel so betrayed and humiliated, he never would have posted all those detailed descriptions of meals consisting of boiled, broiled, braised, baked, grilled, poached, sautéed and stewed crow on the site where she documented her meals. He never would have posted exhaustive, fawning reviews of every film in the *Passion of the Zombie Vampire* heptalogy on the site where she logged her media consumption. He never would have replaced all her profile photos with selections from the ratemylooks.com all-time bottom-20 and updated her relationship status to 'putrescing'.

Doing it, he'd been gleeful and self-satisfied, even euphoric — but, afterwards, he'd felt even worse than before, as if anticipating the price he'd pay. Was now paying.

With cybercrime overtaking fleshcrime in annual costs; with cybercriminals capable, through self-replicating agent programs, of being in an unlimited number of places at once; with prisons overflowing, the Cybercrime Act of 2032 had been overdetermined. The digitization of life that had made cyberbanishment meaningful had also made it inevitable.

Cyberbanishment: the deletion of all his cyberaccounts and, far worse, the deactivation of his arfid and his arfreader. Sitting there, in the park, he was surrounded by people, but, to him, they were faceless ciphers, with no names, no interests, no relationships, none of the scores of personal details that their arfids were broadcasting for everyone else's arfreaders to project onto their retinas. And, to them, he was an ersatz person, revealing nothing and thus capable of anything. Even the dogs, with their mandated arfids, were fuller members of society than him.

For several hours, Julius had tried to make eye contact with someone, with anyone, but people's eyes just passed over him, as if he were an inanimate part of the landscape.

The most he got, at least from adults, were several double takes as people saw him, then, worried that their arfreader was malfunctioning, looked at someone else, and, reassured, glanced, very quickly, back at him. Dogs and small children with no arfreaders did look at him, but their alarmed guardians quickly steered them away.

He reminded himself that he'd lived with no arfreader until his twelfth birthday, but it was one thing to have never experienced the world's true contextual richness, and a very much worse thing to have lost it.

"It's only for a year," he told himself.

"A whole year!" his despondent self retorted.

In a year, his friends would have forgotten him twice over, half his vocabulary would be out of date, the entire culture would have passed him by.

He closed his eyes, and squeezed his fingers into fists. As he tried to recapture, in his mind, the world he'd lost, something wet brushed over his left wrist.

He opened his eyes — a dog, a golden retriever, was licking the back of his hand. Seized by conflicting impulses of gratitude to the dog for acknowledging him and alarm at being so unceremoniously licked by a strange animal, Julius froze.

"Don't worry — he only does that to people he likes," said a woman's voice.

Approaching, looking straight at him: short, athletic, dark-skinned, darker-haired — perhaps Indian — and … that was all he knew about her, which is to say he knew nothing. He couldn't remember the last time he'd conversed with a complete stranger.

Disarming anxiety bloomed in the pit of his stomach. He fought

it back, aware of the irony, and forced himself to focus on her, now just an arm's length away, with his unaugmented sight. Her face was freckled, and pretty.

"I'm Rometa," she said, and offered her hand.

"I'm Julius." As he extended his hand, he was painfully aware that his palm was soaked with sweat.

When they shook, only the softest alarm went off in the back of his head — he didn't know her vaccination status. Then again, her dog had already licked him.

"You must really love dogs," she said.

"Why do you say that?"

"Well, you didn't recoil from him even though he's …" she glanced around, then whispered, "illegal." When she saw the confusion on his face, she raised her eyebrows and added: "No arfid."

He smiled, tasting bitterness.

"That makes two of us," he said.

Now it was her turn to look confused.

"I've been, um, cyberbanished," he said. "Couldn't you tell?"

"I keep my arfreader off most of the time," she said. "Banished, huh? Must be quite a story."

"You keep it off? By choice?"

"To you, that must seem like choosing to live in a prison."

"How do you interact, with people?"

"I talk, I listen, I pay attention. If I want to know something about someone, I just ask."

"But how do you know if you can trust people?"

"I've learned that I usually can. Besides, Shakespeare here is a better judge of character than any arfreader will ever be. Aren't you, boy?"

She bent down and ran her hand over the dog's neck and shoulders, then back around to its throat and chin. The dog smiled and licked her open palm.

"I just can't believe it," Julius said, shaking his head.

"Come on our walk," Rometa said, "and I'll show you a thing or two about this prison called real life."

As he jogged after her into the great unknown mysterious world, Julius felt something he hadn't in a long time. Anticipation. ∎

Igor Teper *lives with his wife and son in the San Francisco Bay Area and teaches old atoms new tricks at temperatures near absolute zero. He also writes stories, occasionally.*

LETTER

A close nuclear black–hole pair in the spiral galaxy NGC 3393

G. Fabbiano[1], Junfeng Wang[1], M. Elvis[1] & G. Risaliti[1,2]

The current picture of galaxy evolution[1] advocates co-evolution of galaxies and their nuclear massive black holes, through accretion and galactic merging. Pairs of quasars, each with a massive black hole at the centre of its galaxy, have separations of 6,000 to 300,000 light years (refs 2 and 3; 1 parsec = 3.26 light years) and exemplify the first stages of this gravitational interaction. The final stages of the black-hole merging process, through binary black holes and final collapse into a single black hole with gravitational wave emission, are consistent with the sub-light-year separation inferred from the optical spectra[4] and light-variability[5] of two such quasars. The double active nuclei of a few nearby galaxies with disrupted morphology and intense star formation (such as NGC 6240 with a separation[6] of about 2,600 light years and Mrk 463 with a separation[7] of about 13,000 light years between the nuclei) demonstrate the importance of major mergers of equal-mass spiral galaxies in this evolution; such mergers lead to an elliptical galaxy[8], as in the case of the double-radio-nucleus elliptical galaxy 0402+379 (with a separation of about 24 light years between the nuclei)[9]. Minor mergers of a spiral galaxy with a smaller companion should be a more common occurrence, evolving into spiral galaxies with active massive black-hole pairs[10], but have hitherto not been seen. Here we report the presence of two active massive black holes, separated by about 490 light years, in the Seyfert[11] galaxy NGC 3393 (50 Mpc, about 160 million light years). The regular spiral morphology and predominantly old circum-nuclear stellar population[12] of this galaxy, and the closeness of the black holes embedded in the bulge, provide a hitherto missing observational point to the study of galaxy/black hole evolution. Comparison of our observations with current theoretical models of mergers suggests that they are the result of minor merger evolution[10].

NGC 3393 (Table 1) was observed with the Chandra X-ray Observatory's camera ACIS-S on 28 February 2004 (ObsID 4868 for 29.7 kiloseconds) and 12 March 2011 (ObsID 12290 for 70 kiloseconds), giving us a total of 89.7 kiloseconds after screening for background events exceeding three standard deviations over the mean background level. We used sub-pixel imaging with a quarter of the native 0.492″ ACIS-S pixel, to recover the mirror resolution (~0.4″ half-maximum radius). We improved the Chandra and Hubble Space Telescope astrometry using stars in the field. Details are in the Supplementary Information. We used XSPEC software for the spectral analysis, and CIAO and DS9 software for other analyses.

Figure 1 shows a marginally extended source in the 3–8-keV band, suggesting some complexity in the emission. The spectrum shows a featureless 3–6-keV continuum and a prominent 6.4-keV Fe K emission line, as previously reported[13]. We find that the continuum and line images differ: there is a single point-like source (upper left of the image, to the northeast, NE) in the 3–6-keV band, coincident (within error) with the nucleus observed by the Hubble Space Telescope, whereas the 6–7-keV image contains two sources with centroids 0.6″ (~150 parsec, pc) apart. The ratio of the 6–7-keV and 3–6-keV images shows relatively more prominent Fe K emission in the source in the lower right of the image (to the southwest, SW), with a position consistent with the maser[14] located by very long baseline interferometry (VLBI). The spectra of both sources (Fig. 2) are typical of Compton-thick active galactic nuclei[13]. Fits to power-law and reflection-component models (Table 2) show that both have observed (2–10 keV) luminosity of a few 10^{40} erg s^{-1}. The emitted luminosities estimated from the Fe K line intensity[13] are 3.4×10^{42} erg s^{-1} (NE) and 5.0×10^{42} erg s^{-1} (SW). These high luminosities and spectral shapes exclude a starburst contribution, consistent with the predominantly old central stellar population[12].

If the 2–10-keV emission were solely from reflected nuclear emission in a Compton-thick active galactic nucleus, the large dimensions of the reflector (thick clouds near the nucleus) would preclude variability. However, variability is implied by the lower flux found in the

Table 1 | Summary of published results from X-ray observations of NGC 3393

Observatory and date of observation	Γ	N_H (cm^{-2})	X-ray band (keV)	Observed X-ray flux, F_X (10^{-13} erg cm^{-2} s^{-1})	X-ray luminosity, L_X (10^{40} erg s^{-1})	Intensity, I(Fe K) (10^{-6} photons cm^{-2} s^{-1})	Equivalent width in Fe K (keV)
BeppoSAX[21] 1 August 1997	1.7 (fixed)	$>1 \times 10^{25}$	20–100	54	180	9.6	$1.9^{+3.8}_{-1.2}$
BeppoSAX[15] 1 August 1997	$2.8^{+1.2}_{-0.7}$	$4.4^{+2.5}_{-1.1} \times 10^{24}$	20–100	NA	NA	14	4 ± 2
XMM-Newton[15] 5 July 2003	1.6 ± 1.2	$>9 \times 10^{23}$	2–10	$0.9^{+0.6}_{-0.4}$	$3.1^{+2.1}_{-1.5}$	2.5	1.4 ± 0.8
Chandra[13] 28 February 2004	1.9 (fixed)	4.7×10^{20} (Gal.)	2–10	2.1 ± 0.04	7.5 ± 0.1(observed) 720 (Fe K) >100 (PEXRAV)	4.2	1.4 ± 0.7
Suzaku[26,27] 23 May 2007	$1.52^{+0.39}_{-0.38}$	$1.7^{+1.4}_{-0.3} \times 10^{24}$	2–10	4	14	4.3	0.5 ± 0.2
Suzaku[26,27] 23 May 2007	$1.52^{+0.39}_{-0.38}$	$1.7^{+1.4}_{-0.3} \times 10^{24}$	15–50	200	680	4.3	0.5 ± 0.2
Swift[28] monitoring	$1.68^{+0.30}_{-0.28}$	NA	14–195	255	890		

Equivalent width is the width of continuum spectrum corresponding to the line flux; Γ is the photon spectral index in a power-law model; N_H is the neutral hydrogen absorption column. 'Gal.' is the Galactic line of sight (absorption due to the Milky Way); 'fixed' means that the variable was not allowed to vary in the fit; NA, not available. The galaxy NGC 3393 (distance about 50 Mpc; $L_H \approx 3.4 \times 10^{10}$ solar luminosities) has a Seyfert 2 nucleus[11], detected in the Hα emission line[17], 13-μm infrared[24] and radio[14,25]. The prominent 6.4-keV Fe Kα line and luminous ($L_X \approx 2 \times 10^{42}$ erg s^{-1}) high-energy X-ray (>10 keV) emission[13,15,21,26–28] suggest a Compton-thick active galactic nucleus. The X-ray luminosity values L_X are observed values including continuum and Fe K line emission, except when inferred from the Fe K or reflection component (using the PEXRAV model[30]) luminosities and noted as such[13].

[1]Harvard-Smithsonian Center for Astrophysics (CfA); 60 Garden Street Cambridge, Massachusetts 02138 USA. [2]INAF-Arcetri; Largo Enrico Fermi 5, Firenze 50125, Italy.
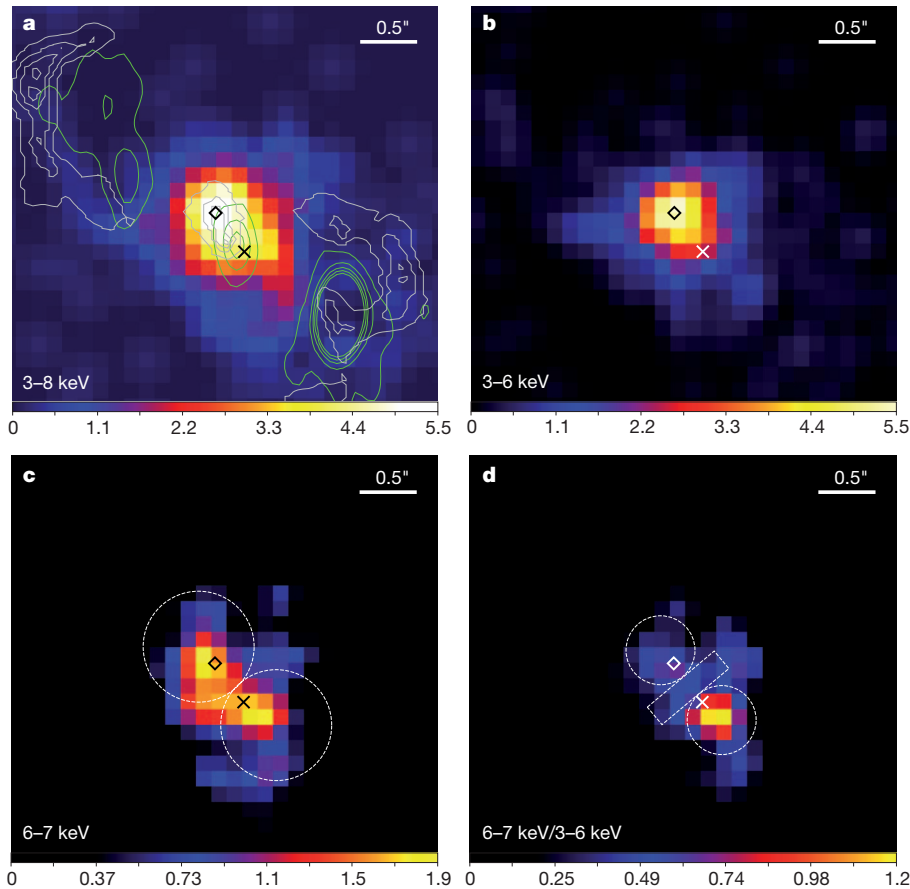
Figure 1 | Chandra ACIS-S images. North is up and East to the left. **a**, Image in the 3–8-keV spectral band of the NGC 3393 nuclear region with quarter-subpixel binning, smoothed with a $\sigma = 0.25''$ Gaussian. Contours of the Hubble Space Telescope F664N Hα emission (ref. 17) and Very Large Array (VLA) radio telescope 8.4-GHz emission (ref. 25) are shown in grey and green, respectively. The diamond and cross indicate the positions of the sources observed by the Hubble Space Telescope[17] and VLBI[14], respectively. The X-ray source contains 279 ± 16 counts in the 3–8-keV band. **b**, Image in the 3–6-keV spectral band, showing continuum emission dominated by the NE source at right ascension (RA) 10 h 48 min 23.47 s and declination (Dec.) $-25^\circ\,09'\,43.1''$ (all positions are J2000.0), coincident with the Hubble Space Telescope position[17]. This position is also consistent with that of the 13-μm source detected in the Very Large Telescope's visible+infrared (VLT/VISIR) image[24]

(P. Gandhi, personal communication). **c**, Image in the 6–7-keV band including both continuum and Fe K line emission; the dashed circles are the spectral counts extraction areas. **d**, Ratio of images shown in **b** and **c**; the values of the ratios in the regions indicated by the dashed outlines are 0.61 ± 0.04 (NW), 1.14 ± 0.10 (SE) and 0.46 ± 0.02 (in between). The Fe K emission is relatively more prominent in a source to the SW of the continuum source, at RA 10 h 48 min 23.45 s, Dec. $-25^\circ\,09'\,43.6''$. The latter position is closer to the nuclear position from the Chandra ObsID 4868 image[13], which was based on the centroid of the Fe K emission. Given the astrometric uncertainty, the SW source is consistent with the VLBI position of the nuclear maser[14]. The sources are visible in both Chandra observations, although statistics are limited in the first. In the previous Chandra analysis[13], subpixel binning and imaging in separate spectral bands were not pursued.

observation[15] by the X-ray Multi-Mirror Mission (XMM)-Newton orbiting space observatory. Transitions from Compton-thick to Compton-thin have been observed in some active galactic nuclei[16], suggesting temporary 'holes' in the wall of obscuring clouds. The dimming observed by XMM-Newton may be related to the NE source, because an intrinsically obscured power-law component, $N_E \propto E^{-\gamma}$, where E is energy, with parameters $\Gamma \approx 1.9$ and $N_H \approx 2 \times 10^{23}\,\text{cm}^{-2}$ could fit its spectrum, suggesting that some direct nuclear emission may be visible. A passing broad-line-region cloud of $N_H \approx 2 \times 10^{23}\,\text{cm}^{-2}$ may have obscured this component during the XMM-Newton observation, leaving only residual scattered continuum. Even so, given the large Fe K equivalent width (Table 2), we may be seeing only reflected emission. In this case, we would be observing an additional absorption of $N_H \approx 2 \times 10^{23}\,\text{cm}^{-2}$ towards the (warm) reflector.

The SW nuclear source, with weaker, flat continuum, and no optical Hα counterpart[17] (Fig. 1), has prominent, and possibly complex, Fe K emission. Although its luminosity is consistent with the most luminous ultraluminous sources detected in nearby galaxies, its spectrum argues for a Compton-thick active galactic nucleus. A 6.5-keV Fe line has been reported in the M82 X-1 ultraluminous source, but this line is

very broad and its equivalent width is model-dependent[18]. The Fe K line from the SW source instead has the well-defined narrow core found in Compton-thick nuclei, understood to be fluorescent emission excited by the scattered nuclear radiation[19]. Moreover, unlike the SW source, strong continuum emission dominates the Chandra spectra of ultraluminous sources[20].

We can rule out the interpretation that the two sources are the result of a single active galactic nucleus interacting with clouds. Although the spectral shape of the emission of the SW source cannot exclude a local mirror reflecting flux from the NE source, given the $\sim$150-pc separation and the Chandra limit on size of $<0.4''$, the SW source covering factor exceeds 1/70, implying an intrinsic luminosity of the NE source well above its measured value. Also, in the reflection hypothesis, we should be detecting Hα and [O III] emission from the SW source, but none is seen[17]; this lack argues for a totally cocooned source, where the Fe K emission is seen in transmission[13]. The spatial coincidence with the maser emission[14] observed by VLBI (Fig. 1) reinforces this conclusion. The NE source cannot be due to reflection by a similarly small reflector, given both its variability[15] and the modest far-infrared luminosity of the system: $\sim$3 × 10^{43} erg s^{-1} at 60 μm, based on the Infrared Astronomical Satellite (IRAS) point source catalogue flux of 2.25 Jy, a

Figure 2 | X-ray spectra. a, The NE source, with intrinsically absorbed power-law best fit. **b**, The SW source, with best-fit PEXRAV (Table 1) model. Errors are one standard deviation. Spectra were extracted from the regions shown in Fig. 1c with background from a source free region 10″ to the east of the nuclei, and fitted with XSPEC (version 12.6.0) using the C-statistic. The NE spectrum, which shows a downturn around 3 keV, is well fitted with an intrinsically absorbed (in addition to the line-of-sight Galactic $N_H \approx 4.7 \times 10^{20}$ cm^{-2}) power-law continuum plus Fe K line, although the uncertainties in the parameters are large and this model is not statistically a unique choice. Given the flat continuum and more prominent Fe K line, the SW source is likely to be Compton thick. An intrinsically absorbed power-law model fit gives only flat power laws and no intrinsic absorption. We have assumed a power-law + reflection component model for the continuum, fixing the N_H to the Galactic value, that would be representative of Compton-thick emission[13]. This model was also used for the NE source, as an alternative to the intrinsically absorbed power law. In the SW source the luminosity is dominated by the Fe K emission. X-ray luminosity $L_X(3-6 \text{ keV}) = 3.7 \pm 0.7 \times 10^{39}$ erg s^{-1}, whereas $L_X(\text{Fe K}) = 1.2 \pm 0.3 \times 10^{40}$ erg s^{-1}. The equivalent width of the Fe K line is $1.16^{+0.82}_{-0.58}$ keV (NE), and $2.77^{+1.95}_{-1.43}$ keV (SW) at the 68% confidence level. The Fe K line of the SW source appears broadened (0.13 ± 0.05 keV) and complex, with a main peak at Fe Kα and a secondary peak at Fe Kβ. The ratio of the Kβ intensity to the Kα intensity, $I(\text{K}\beta)/I(\text{K}\alpha)$, is 0.35 ± 0.19, consistent with the expected range of $I(\text{K}\beta)/I(\text{K}\alpha) \approx 0.12-0.2$ for neutral or weakly ionized iron[29]. The results are listed in Table 2.

standard accretion-rate-to-luminosity conversion efficiency for quasars of 10% for the Eddington accretion rate, yield masses of $\sim 8 \times 10^5$ solar masses and $\sim 10^6$ solar masses for the NE and SW sources, respectively. However, lower efficiencies and sub-Eddington accretion are possible, so the above masses are lower limits. The dynamic VLBI mass measurement[14] of $\sim 3 \times 10^7$ solar masses for the SW source implies that the product of efficiency and accretion rate must be lower by a factor of approximately 30.

If the masses of the two massive black holes are similar, the relation[23] of the mass of the galactic bulge to the mass of the massive black hole suggests that NGC 3393 may be the remnant of a major merger of two similar spiral galaxies. After about five billion years, this merger would produce a remnant with prominent 'grand-design' spiral arms and massive black-hole binary separation, as seen in NGC 3393 (ref. 8). However, the stellar population of this bulge would be significantly rejuvenated (L. Mayer, personal communication), which is at odds with the age of the stellar population in the central 200 pc of NGC 3393 (ref. 12). Moreover, in a major merger, the merging time-scale[8] for two massive black holes at a separation of 150 pc is about a million years, making the detection of such events rare.

Although this occurrence cannot be excluded on the basis of a single detection, a better explanation of our results may be the merger of galaxies of unequal mass (and therefore of massive black holes of unequal mass), which would result in a longer, billion-year timescale[10], and would be consistent with the lack of widespread star formation. The constraints on the masses of the two massive black holes allow this possibility. Interestingly, minor mergers may also have resulted in the growth of one of the massive black holes by promoting more active nuclear accretion in the smaller massive black hole[17]. The denser

value comparable with the estimated intrinsic nuclear emission. Finally, a jet interacting with the interstellar medium in the host galaxy bulge is also ruled out because the luminosity measured in the Fe K line would require extreme conditions in terms of shock energy and gas ionization, and produce copious soft X-ray emission not seen with Chandra.

We conclude that there are two obscured active galactic nuclei, each powered by accretion onto a massive black hole, in the central regions of NGC 3393. The emitted luminosities inferred from the Fe K lines are both a few 10^{42} erg s^{-1}, showing that both sources contribute to the Compton-thick emission seen by the X-ray astronomy satellite BeppoSAX[21], with the SW source being more prominent.

The inferred intrinsic X-ray luminosities, for the spectral energy distribution (SED) of a standard active galactic nucleus[22], and using a

Table 2 | Results of the spectral analysis of sources NE and SW

Source; counts	Γ	N_H (cm^{-2})	F_X(2–10 keV) (10^{-13} erg cm^{-2} s^{-1})	L_X(2–10 keV) (10^{40} erg s^{-1})	I(Fe K) (10^{-6} photons cm^{-2} s^{-1})	L_X(Fe K) (10^{39} erg s^{-1})	Equivalent width in Fe K (keV)	C-stat. (d.o.f.)
NE; 134 ± 9	$1.9^{+0.6}_{-1.8}$	$(2.1 \pm 0.6) \times 10^{23}$	$1.3^{+0.3}_{-0.1}$	$4.4^{+0.8}_{-0.5}$ (observed) 340 (Fe K)	2.0 ± 0.5	7 ± 1.7	$1.2^{+0.8}_{-0.6}$	263 (337)
NE; 134 ± 9	1.7 (fixed)	4.7×10^{20}(Gal.)	1.2 ± 0.2	3.8 ± 0.6 (observed) >60 (PEXRAV)	1.9 ± 0.5	7 ± 1.7	$1.0^{+0.7}_{-0.5}$	263 (338)
SW; 75 ± 8	1.7 (fixed)	4.7×10^{20}	$0.7^{+0.2}_{-0.1}$	$2.3^{+0.7}_{-0.3}$ (observed) 500 (Fe K) >70 (PEXRAV)	2.6 ± 0.6	12.0 ± 2.7	$2.8^{+1.9}_{-1.4}$	200 (338)

C-stat., statistic from the C method; d.o.f., degrees of freedom. The counts of each source were extracted in the 3–8 keV band. Flux and luminosity were estimated in the 2–10 keV band to compare with results in the literature. Flux and luminosity are observed values including continuum and Fe K line emission, except when inferred from the Fe K or reflection component (PEXRAV) luminosities and so noted, in which case they are estimates of the emitted values. Errors are one standard deviation (68%). Aperture corrections were applied to account for the ~35% of the spectral photons in the wings of the Chandra point-spread function missed from the extraction regions. The total observed (2–10 keV) luminosity of the combined NE+SW emission is consistent with the results from the first Chandra observation[13], which did not resolve the two sources (see Table 1 for these comparisons). Considering the difference in beam size, and the luminosity of the rest of the galaxy[13], these fluxes are also consistent with the measurement from the X-ray astronomy satellite Suzaku. The XMM-Newton measurement instead gives a lower value of the combined luminosity, suggesting a factor-of-2 variability in the 2–10 keV range for the total NE+SW flux over a one-year timescale. At higher energies (>15 keV), the 5-year light curve of NGC 3393 (ref. 28) from the Burst Alert telescope (BAT) on the spacecraft Swift does not show any significant variability.

circum-nuclear environment of the SW source, as suggested by our results, may have resulted from such a process.

1. Colpi, M. & Dotti, M. Massive binary black holes in the cosmic landscape. In *Advanced Science Letters, Special Issue on Computational Astrophysics* (ed. Mayer, L.) http://arxiv.org/abs/0906.4339 (2009).
2. Green, P. J. *et al.* SDSS J1254+0846: a binary quasar caught in the act of merging. *Astrophys. J.* **710,** 1578–1588 (2010).
3. Civano, F. *et al.* A runaway black hole in COSMOS: gravitational wave or slingshot recoil? *Astrophys. J.* **717,** 209–222 (2010).
4. Boroson, T. A. & Lauer, T. R. A candidate sub-parsec supermassive binary black hole system. *Nature* **458,** 53–55 (2009).
5. Valtonen, M. J. *et al.* A massive binary black-hole system in OJ287 and a test of general relativity. *Nature* **452,** 851–853 (2008).
6. Komossa, S. *et al.* Discovery of a binary active galactic nucleus in the ultraluminous infrared galaxy NGC 6240 using Chandra. *Astrophys. J.* **582,** L15–L19 (2003).
7. Hutchings, J. B. & Neff, S. G. The double-nucleus galactic merger MKN 463. *Astron. J.* **97,** 1306–1311 (1989).
8. Mayer, L. *et al.* Rapid formation of supermassive black hole binaries in galaxy mergers with gas. *Science* **316,** 1874–1877 (2007).
9. Rodriguez, C. *et al.* A compact supermassive binary black hole system. *Astrophys. J.* **646,** 49–60 (2006).
10. Callegari, S. *et al.* Growing massive black hole pairs in minor mergers of disk galaxies. *Astrophys. J.* **729,** 85, http://dx.doi.org/ 10.1088/0004-637X/729/2/85 (2011).
11. Diaz, A. I., Prieto, M. A. & Wamsteker, W. The optical and UV spectrum of the Seyfert type 2 galaxy NGC 3393. *Astron. Astrophys.* **195,** 53–59 (1988).
12. Cid Fernandes, R. *et al.* The star formation history of Seyfert 2 nuclei. *Mon. Not. R. Astron. Soc.* **355,** 273–296 (2004).
13. Levenson, N. A., Heckman, T. M., Krolik, J. H., Weaver, K. A. & Życki, P. T. Penetrating the deep cover of Compton-thick active galactic nuclei. *Astrophys. J.* **648,** 111–127 (2006).
14. Kondratko, P. T., Greenhill, L. J. & Moran, J. M. The parsec-scale accretion disk in NGC 3393. *Astrophys. J.* **678,** 87–95 (2008).
15. Guainazzi, M., Fabian, A. C., Iwasawa, K., Matt, G. & Fiore, F. On the transmission-dominated to reprocessing-dominated spectral state transitions in Seyfert 2 galaxies. *Mon. Not. R. Astron. Soc.* **356,** 295–308 (2005).
16. Risaliti, G. *et al.* The XMM-Newton long look of NGC 1365: uncovering of the obscured X-ray source. *Mon. Not. R. Astron. Soc.* **393,** L1–L5 (2009).
17. Cooke, A. J. *et al.* The narrow-line region in the Seyfert 2 galaxy NGC 3393. *Astrophys. J. Suppl. Ser.* **129,** 517–545 (2000).
18. Strohmayer, T. E. & Mushotzky, R. F. Discovery of X-ray quasi-periodic oscillations from an ultraluminous x-ray source in M82: evidence against beaming. *Astrophys. J.* **586,** L61–L64 (2003).
19. George, I. M. & Fabian, A. C. X-ray reflection from cold matter in active galactic nuclei and X-ray binaries. *Mon. Not. R. Astron. Soc.* **249,** 352–367 (1991).
20. Fabbiano, G. *et al.* The time-variable ultraluminous X-ray sources of ''The Antennae''. *Astrophys. J.* **584,** L5–L8 (2003).
21. Maiolino, R. *et al.* Heavy obscuration in X-ray weak AGNs. *Astron. Astrophys.* **338,** 781–794 (1998).
22. Elvis, M. *et al.* Atlas of quasar energy distributions. *Astrophys. J. Suppl. Ser.* **95,** 1–68 (1994).
23. Magorrian, J. *et al.* The demography of massive dark objects in galaxy centers. *Astron. J.* **115,** 2285–2305 (1998).
24. Gandhi, P. *et al.* Resolving the mid-infrared cores of local Seyferts. *Astron. Astrophys.* **502,** 457–472 (2009).
25. Schmitt, H. R., Ulvestad, J. S., Antonucci, R. R. J. & Kinney, A. L. Jet directions in Seyfert galaxies: radio continuum imaging data. *Astrophys. J. Suppl. Ser.* **132,** 199–209 (2001).
26. Miyazawa, G. T. Haba, Y. & Kunieda, H. Broad-band temporal and spectral variation of 36 active galactic nuclei observed with Suzaku. *Publ. Astron. Soc. Jpn* **61,** 1331–1354 (2009).
27. Fukazawa, Y. *et al.* Fe-K line probing of material around the active galactic nucleus central engine with Suzaku. *Astrophys. J.* **727,** 19–31 (2011).
28. Cusumano, G. *et al.* The Palermo Swift-BAT hard X-ray catalogue. III. Results after 54 months of sky survey. *Astron. Astrophys.* **524,** A64–A101 (2010).
29. Palmeri, P., Mendoza, C., Kallman, T. R., Bautista, M. A. & Meléndez, M. Modeling of iron K lines: radiative and Auger decay data for Fe II-Fe IX. *Astron. Astrophys.* **410,** 359–364 (2003).
30. Magdziarz, P. & Zdziarski, A. A. Angle-dependent Compton reflection of X-rays and gamma-rays. *Mon. Not. R. Astron. Soc.* **273,** 837–848 (1995).

Structural basis of PIP$_2$ activation of the classical inward rectifier K$^+$ channel Kir2.2

Scott B. Hansen[1], Xiao Tao[1] & Roderick MacKinnon[1]

The regulation of ion channel activity by specific lipid molecules is widely recognized as an integral component of electrical signalling in cells[1,2]. In particular, phosphatidylinositol 4,5-bisphosphate (PIP$_2$), a minor yet dynamic phospholipid component of cell membranes, is known to regulate many different ion channels[2–8]. PIP$_2$ is the primary agonist for classical inward rectifier (Kir2) channels, through which this lipid can regulate a cell's resting membrane potential[2,7–9]. However, the molecular mechanism by which PIP$_2$ exerts its action is unknown. Here we present the X-ray crystal structure of a Kir2.2 channel in complex with a short-chain (dioctanoyl) derivative of PIP$_2$. We found that PIP$_2$ binds at an interface between the transmembrane domain (TMD) and the cytoplasmic domain (CTD). The PIP$_2$-binding site consists of a conserved non-specific phospholipid-binding region in the TMD and a specific phosphatidylinositol-binding region in the CTD. On PIP$_2$ binding, a flexible expansion linker contracts to a compact helical structure, the CTD translates 6 Å and becomes tethered to the TMD and the inner helix gate begins to open. In contrast, the small anionic lipid dioctanoyl

glycerol pyrophosphatidic acid (PPA) also binds to the non-specific TMD region, but not to the specific phosphatidylinositol region, and thus fails to engage the CTD or open the channel. Our results show how PIP$_2$ can control the resting membrane potential through a specific ion-channel-receptor–ligand interaction that brings about a large conformational change, analogous to neurotransmitter activation of ion channels at synapses.

PIP$_2$ influences the metabolic state of cells by at least three distinct pathways (Supplementary Fig. 1a, b): first, as the prototypical second messenger being cleaved into diacyl glycerol and inositol triphosphate[10,11]; second, as a localization signal targeting soluble proteins to the plasma membrane[12–14]; and third, as a signalling molecule capable of agonizing an ion channel[2–8,15,16]. This latter role, in which an ion channel is activated by PIP$_2$, was first discovered in 1998 when it was shown that PIP$_2$ acted alone to open a Kir channel[8].

Figure 1a, b shows the influence of PIP$_2$ on the function of Kir2.2 from chicken. Following excision of an inside-out membrane patch from a *Xenopus* oocyte expressing Kir2.2 channels, initially large
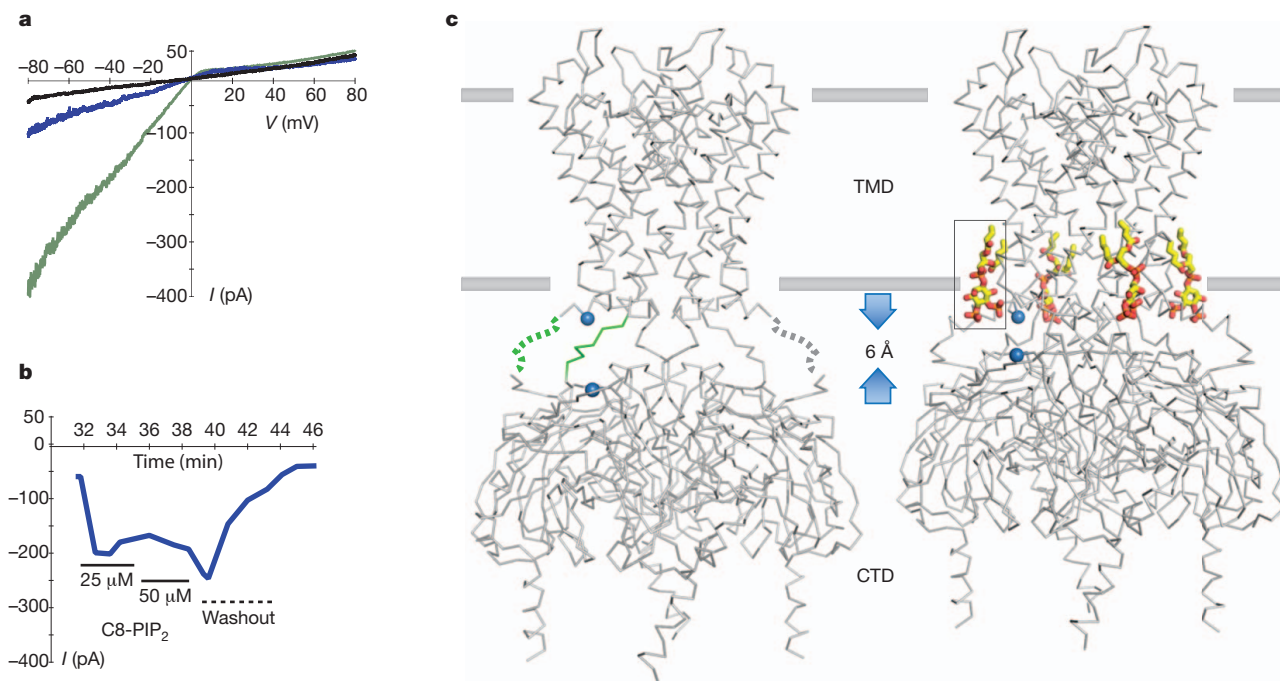
Figure 1 | Effect of a short-chain PIP$_2$ on Kir2.2. a, Endogenous PIP$_2$ depletion causes 'run down' of Kir2.2 channels in an excised inside-out patch from *Xenopus* oocytes as shown by the three macroscopic current traces recorded with a voltage ramp from −80 to +80 mV immediately (green), 30 min (blue) and 50 min (black) after patch excision. **b,** The short-chain PIP$_2$ added to the bath solution (solid line with concentration indicated below) beginning 32 min after patch excision partially rescued Kir2.2 channel activity. The bath was then perfused (dashed line) at time = 40 min with ~1 ml min^{-1} bath solution for 3 min. **c,** X-ray crystal structures of apo- (left, PDB code 3JYC) and PIP$_2$-bound (right, PDB code 3SPI) Kir2.2

tetramer (grey α-carbon traces) viewed from the side with the extracellular solution above. The lipid bilayer boundaries are shown as grey bars. Four PIP$_2$ molecules are shown as sticks and coloured according to atom type: carbon, yellow; phosphorous, orange; and oxygen, red. One PIP$_2$ molecule in a similar orientation as in Fig. 2a is outlined by a black box. On PIP$_2$ binding the flexible linker between CTD and TMD consisting of two strands (highlighted green for one subunit, dotted line indicating disordered region in the crystal structure) form helical structures, and the CTD translates towards the TMD by 6 Å. A set of reference atoms (Asp 72 and Lys 220 α-carbons) are highlighted as blue spheres in each structure.

[1]Laboratory of Molecular Neurobiology & Biophysics, The Rockefeller University, Howard Hughes Medical Institute, 1230 York Avenue, New York, New York 10065, USA.

inward K$^+$ currents diminish over time. The diminution occurs because PIP$_2$ is depleted from the membrane's inner leaflet[8]. The K$^+$ currents can be restored partially by exposing the cytoplasmic face of the patch to the short-chain derivative of PIP$_2$ in a dose-dependent manner[17,18] (Fig. 1b). PIP$_2$ is the primary agonist for Kir2 channels, through which this lipid can regulate a cell's resting membrane potential. Here we use X-ray crystallography to understand the mechanism by which PIP$_2$ opens a Kir2 channel.

Kir2.2 is a tetrameric ion channel comprised of a TMD, which forms the prototypic potassium-selective pore, and a large CTD, which characterizes all Kir channels[19] (Fig. 1c). We determined the structures of wild-type Kir2.2 from chicken (with disordered segments of the amino and carboxy termini truncated) in the presence of the short-chain derivative of PIP$_2$ at 3.3 Å resolution (Fig. 1c). We also determined the structures of two point mutants, I223L and R186A, in the presence of PIP$_2$ at 3.0 Å and 2.6 Å resolutions, respectively (Supplementary Table 1). These mutants were studied because they are described in the literature as altering the apparent affinity for PIP$_2$[9,20]. All three channels have overall similar structures and taken together enhance our knowledge of the detailed chemical properties through which PIP$_2$ binds to and modifies the channel's structure and function (Supplementary Fig. 2a, b). Electron density maps are of high quality for the entire protein, and strong density for the three phosphates in PIP$_2$—observed in all three structures—allowed accurate placement of the ligand (Supplementary Fig. 3a). Furthermore, the glycerol backbone of PIP$_2$ is well ordered and easily placed in the higher resolution structures, along with 4–6 carbons of the lipid acyl chains (Supplementary Fig. 3a). One PIP$_2$ molecule binds to each of the four channel subunits (Fig. 1c).

PIP$_2$ binds at the interface between the TMD and CTD and produces a large conformational change in Kir2.2 (Fig. 1c). The entire CTD translates 6 Å towards the TMD in association with the formation of two new helices, an N-terminal extension of the 'interfacial' helix and a 'tether' helix at the C terminus of the inner helix (Fig. 2a and Supplementary Fig. 3b). The 6 Å translation of the CTD is reflected in a compression along the c-axis of the unit cell (Supplementary Table 1). The protein conformational changes position amino acids that form the binding site for the 4′,5′-phosphate-substituted inositol head group of PIP$_2$.

The PIP$_2$-binding site comprises amino acids from two main structural regions of the channel. The acyl chains, glycerol backbone and 1′ (phosphodiester) phosphate of PIP$_2$ interact with the TMD, while the inositol head group makes interactions with the CTD (Fig. 2a and Supplementary Fig. 4). In detail, the acyl chains insert into the membrane layer where they interact with hydrophobic amino acids on both the inner and outer helices, while the 1′ phosphate makes interactions with amino acids forming the sequence arginine-tryptophan-arginine (amino acids 78–80 in Kir2.2) (Fig. 2b). This sequence is conserved as arginine-tryptophan-arginine or lysine-tryptophan-arginine among many different Kir channels, and the reason for this conservation is made clear by the PIP$_2$ complex: the arginine-tryptophan-arginine sequence is located at the N terminus of the outer helix and forms a binding site in which the 1′ phosphate caps the helix and is cradled by main-chain amide nitrogen atoms and the guanidinium groups from the two arginine residues (Fig. 2a and Supplementary Fig. 4). The tryptophan residue appears to anchor the end of the outer helix at the membrane interface and also interact with one of the acyl chains. With the acyl chains, glycerol backbone and 1′ phosphate of the lipid molecule contacting the TMD, the inositol ring of the head group is oriented towards the CTD, where the 4′ and 5′ phosphates are positioned to interact directly with Lys 183, Arg 186, Lys 188 and Lys 189 (Fig. 2a and Supplementary Fig. 4). The latter two positively charged amino acids are located on the tether helix, the structure of which is induced by the binding of PIP$_2$. Other amino acids on the tether helix, including Arg 190, participate in the formation of a hydrogen-bonding network that seems to strengthen the interaction between the tether
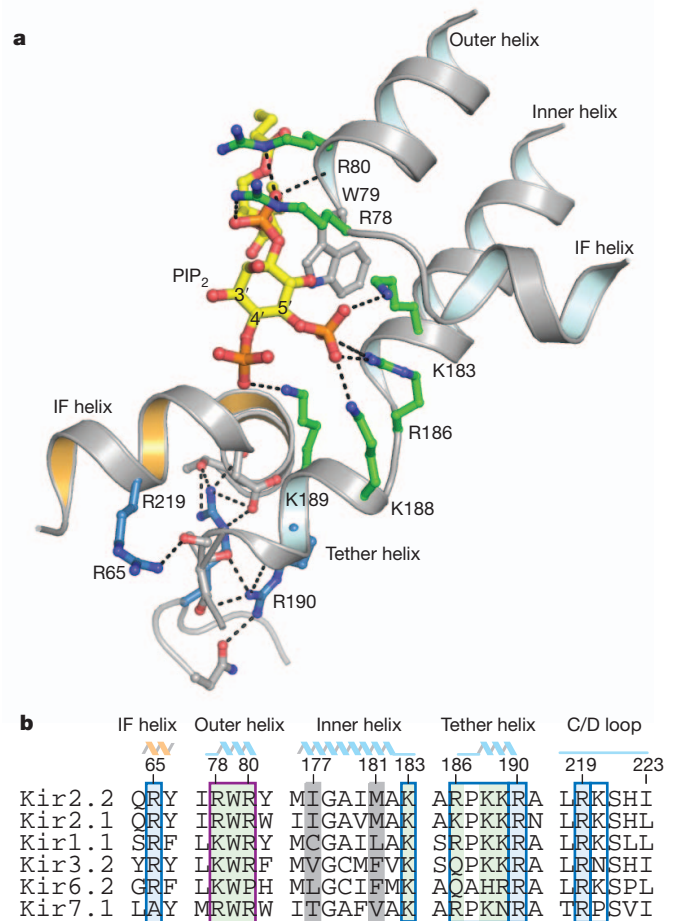
Figure 2 | PIP$_2$-binding site. a, A detailed view of the PIP$_2$-binding site is shown in a similar orientation as outlined in Fig. 1c. Helices (shown as ribbon) from different subunits are distinguished by their interior colour (orange and cyan). Residues hydrogen bonded (dashed lines) to PIP$_2$ are coloured green, and residues stabilizing the PIP$_2$-binding site in the CTD but lacking direct contact are coloured blue. All side chains are shown as sticks. PIP$_2$ is shown as sticks and coloured according to atom type: carbon, yellow; phosphorous, orange; and oxygen, red. **b,** An amino acid sequence alignment of selected eukaryotic Kir channels showing residues predicted from the literature (blue outline) and not predicted (purple outline) to interact with PIP$_2$[9,28]. Residues with direct bonding interactions to PIP$_2$ and with a structural role are highlighted in green and blue, respectively. The two residues serving as the inner helix gate are highlighted in grey.

helix and other regions of the CTD, especially the N-terminal extension of the interfacial helix, the structure of which is also induced by the binding of PIP$_2$ (Fig. 2a and Supplementary Fig. 4). A sequence alignment shows that the amino acids binding to PIP$_2$ are highly conserved among the large family of inward rectifier K$^+$ channels (Fig. 2b). Because all members of this ion channel family seem to be regulated by PIP$_2$ (some in concert with other ligands such as ATP or G proteins)[16], we anticipate that the PIP$_2$ site described here will be observed in many other inward rectifiers.

The detailed chemical properties of the PIP$_2$-binding site suggest that the TMD region should bind to any lipid that contains a glycerol backbone, acyl chains and a 1′ phosphate, whereas the CTD should provide the specificity for the inositol phosphate head group. Moreover, because the head group region of the binding site is formed only after the conformational changes occur in the channel, we would predict that a glycerol phospholipid without an inositol head group would bind to the TMD but not induce the conformational changes. We tested this prediction by determining a 2.45 Å resolution crystal structure of Kir2.2 in the presence of a short-chain derivative of pyrophosphatidic acid
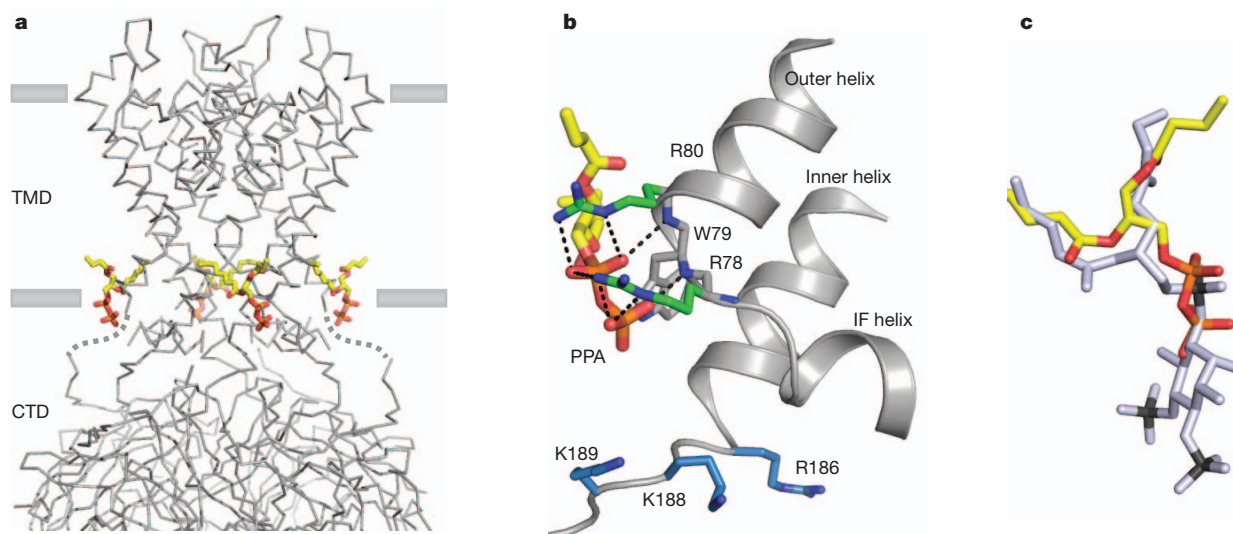
Figure 3 | Conserved non-specific lipid-binding site in Kir channels. **a**, A grey α-carbon representation of Kir2.2 tetramer in complex with PPA, a small anionic lipid lacking an inositol ring. PPA-bound Kir2.2 assumes a closed conformation similar to apo-Kir2.2 (PDB code 3JYC) with the flexible linker elongated and the CTD unengaged. The four PPA molecules are shown as sticks and coloured according to atom type: carbon, yellow; phosphorous, orange; and oxygen, red. **b**, A close-up view of the PPA-binding site. PPA contacts Kir at the cytoplasmic end of the outer helix making strong interactions with the guanidiniums of R78 and R80 and the backbone amide nitrogens of the helix turn; similar to the interactions of the 1′ phosphate of PIP$_2$. However, residues (blue sticks) for interacting with the PIP$_2$ inositol-4′,5′-phosphate remain distant to the lipid-binding site; R186 orients with its side chain pointing towards the ion-conduction pathway. **c**, Superposition of PPA (coloured the same as in **a**) and PIP$_2$ (grey).

(PPA), which contains as a head group only phosphoric acid instead of the 4′,5′-phosphate-substituted inositol ring (Supplementary Table 1). This lipid is bound to the TMD in a manner almost identical to PIP$_2$; however, the head group does not interact with the CTD and the protein conformational changes induced by PIP$_2$ do not occur (Fig. 3a–c). This finding is compatible with recent functional studies showing that small head group anionic lipids failed to activate Kir channels in the absence of PIP$_2$[21].

We wish to understand how the PIP$_2$-induced conformational changes relate to channel activity. Comparison of the inner helix gate in the PIP$_2$ and PPA complexes shows that the interaction of the CTD with the TMD induced by PIP$_2$ is key to opening the gate (Fig. 4a–c). In the PPA complex, in which the CTD is extended away from the TMD, the gate in the TMD is tightly closed (4.9 Å at Ile 177), whereas in the PIP$_2$ complex the inner helices have begun to separate (6.3 Å). The separation of helices comes about as a result of a slight splaying, but more significantly a rotation of the inner helices, which moves hydrophobic amino acid side chains away from the ion pathway (Fig. 4a–c). Opening of the inner helix gate to approximately 6.3 Å (approximately 5 Å diameter between van der Waals surfaces of carbon atoms at the narrowest region) is probably still insufficient to permit ion conduction, but the gate is clearly on the way to an open conformation. A previously published study of prokaryotic Kir channels proposed that interactions between the TMD and CTD of those channels influence the distribution of ions in the selectivity filter[22]. We observe no such influence of the CTD on ions in the filter in the high-quality electron density maps in our analysis of Kir2.2 (Supplementary Fig. 5a–c). In the present study, the data support a simple allosteric mechanism of gating control by the signalling lipid PIP$_2$, in which the lipid mediates docking of the CTD to the TMD and opening of the inner helix gate, as depicted in Fig. 4d.

The ion pathway in Kir channels has a second constriction formed by the G-loop, at the apex of the CTD. This loop in some instances is thought to function as a gate, referred to as the G-loop gate[23]. In Kir2.2 the conformation of the G-loop is altered by PIP$_2$, either directly through the binding of PIP$_2$ or indirectly through the docking of the CTD to the TMD when PIP$_2$ binds (Supplementary Fig. 6a, b). But it seems that PIP$_2$ does not control the G-loop gate to a large extent, because in both conformations this gate is open (smallest diameter 7.8 Å) (Supplementary Fig. 6a). The mutation I223L affects the

conformation of the G-loop gate in a manner that might explain the apparent increased affinity for PIP$_2$[20]. In this mutant, although the CTD does not bind to the TMD and the tether and interfacial helices do not form, the G-loop adopts its PIP$_2$-bound conformation (Supplementary Fig. 6c, d). It is thus possible that the mutant favours PIP$_2$
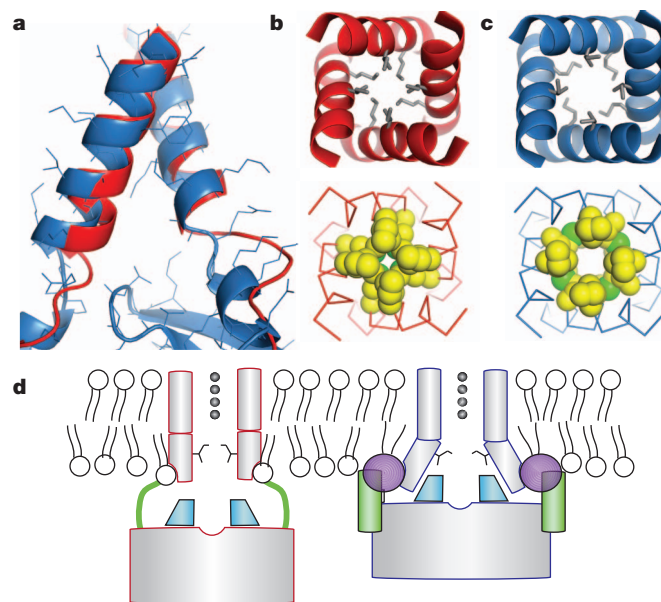
Figure 4 | A proposed mechanism of Kir2.2 activation by PIP$_2$.
a, Superposition of the TMD inner helices of the PIP$_2$-bound (blue ribbon) and apo- (red ribbon) Kir2.2 structures. PIP$_2$ binding results in a splaying of the helices near the helix bundle activation gate. **b**, **c**, Comparison of the inner helix bundle gate in PPA-bound Kir2.2 (**b**) and PIP$_2$-bound Kir2.2 (**c**) viewed from the extracellular side. Side chains of the residues in the bundle crossing are represented as either grey sticks or space-filling CPK models (carbon, yellow; and sulphur, green). **d**, A proposed mechanism for Kir2.2 activation by PIP$_2$. PIP$_2$ (purple sphere) binds at an interface between the TMD (grey cylinder) and the CTD (grey rectangle) and induces a large conformational change: a flexible linker (green line) contracts to a compact helical structure (green cylinder), the CTD translates towards and becomes tethered to the TMD, the G-loop (cyan wedge) inserts into the TMD and the inner helix activation gate opens.

binding by tending to favour the bound configuration before PIP_2 binds.

The membrane lipid PIP_2 has a central role in cell signalling through three distinct pathways (Supplementary Fig. 1a). In one of these pathways PIP_2 acts directly on specific ion channels to regulate their activity. PIP_2 is the primary agonist for Kir2 channels, which control the resting membrane potential in many cells. Since this discovery more than ten years ago, this form of ion channel regulation has been a topic of intense study. The crystal structures presented here reveal the mechanism of PIP_2 activation of Kir2 channels. PIP_2 binds to a lipid-binding site at the membrane's inner leaflet, and through specific interactions between the 4′,5′-inositol-phosphate head group and the channel a large conformational change occurs, initiating pore opening.

METHODS SUMMARY

Chicken Kir2.2 with a C-terminal GFP and a 1D4 epitope was expressed in *Pichia* and purified in n-decyl-β-D-maltopyranoside (DM, Anatrace) by 1D4 antibody affinity chromatography followed by PreScission protease cleavage and gel filtration[19]. Purified protein was concentrated to $9\,mg\,ml^{-1}$ and mixed with freshly prepared dioctanoyl PIP_2 (10 mM stock in water) or dioctanoyl PPA (100 mM stock in water) to a final concentration of 0.6–1 mM and 5 mM, respectively. Crystals, diffracting between 2.45 and 3.3 Å, were obtained from a 200 nl hanging drop with 4 mM DM, 20 mM dithiothreitol, 3 mM TCEP, 0.5 M KCl and PEG 400 or PEG 4000 as a precipitant and cryoprotected in reservoir solution containing 25–30% glycerol. Phases were obtained by molecular replacement with apo-Kir2.2 (PDB code 3JYC) using MolRep[24] in the CCP4 suite[25]. The models were built in Coot[26] and refined in Phenix[27] to R_{free} of 0.22 to 0.28. Complete crystallographic data and refinement statistics are shown in Supplementary Information.

Electrophysiology experiments were conducted using patch clamp on *Xenopus* oocytes expressing wild-type Kir2.2. Briefly, oocytes were injected with 50 nl ($\sim 2\,mg\,ml^{-1}$) cRNA and used for patch recording after 2–3 days. Large pipette tips with typical resistance of 0.4–0.9 MΩ were used.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Dart, C. Lipid microdomains and the regulation of ion channel function. *J. Physiol. (Lond.)* **588,** 3169–3178 (2010).
2. Hilgemann, D. W., Feng, S. & Nasuhoglu, C. The complex and intriguing lives of PIP2 with ion channels and transporters. *Sci. STKE* **2001,** re19 (2001).
3. Suh, B. C. & Hille, B. Regulation of ion channels by phosphatidylinositol 4,5-bisphosphate. *Curr. Opin. Neurobiol.* **15,** 370–378 (2005).
4. Fujiwara, Y. & Kubo, Y. Regulation of the desensitization and ion selectivity of ATP-gated P2X2 channels by phosphoinositides. *J. Physiol. (Lond.)* **576,** 135–149 (2006).
5. Logothetis, D. E., Jin, T., Lupyan, D. & Rosenhouse-Dantsker, A. Phosphoinositide-mediated gating of inwardly rectifying K^+ channels. *Pflügers Arch. Eur. J. Physiol.* **455,** 83–95 (2007).
6. Vaithianathan, T. *et al.* Direct regulation of BK channels by phosphatidylinositol 4,5-bisphosphate as a novel signaling pathway. *J. Gen. Physiol.* **132,** 13–28 (2008).
7. Gamper, N. & Shapiro, M. S. Regulation of ion transport proteins by membrane phosphoinositides. *Nature Rev. Neurosci.* **8,** 921–934 (2007).
8. Huang, C. L., Feng, S. & Hilgemann, D. W. Direct activation of inward rectifier potassium channels by PIP2 and its stabilization by Gβγ. *Nature* **391,** 803–806 (1998).
9. Lopes, C. M. B. *et al.* Alterations in conserved Kir channel-PIP2 interactions underlie channelopathies. *Neuron* **34,** 933–944 (2002).
10. Berridge, M. J. Inositol trisphosphate and calcium signalling. *Nature* **361,** 315–325 (1993).
11. Monserrate, J. P. & York, J. D. Inositol phosphate synthesis and the nuclear processes they affect. *Curr. Opin. Cell Biol.* **22,** 365–373 (2010).
12. Martin, T. F. PI(4,5)P2 regulation of surface membrane traffic. *Curr. Opin. Cell Biol.* **13,** 493–499 (2001).
13. Cho, W. & Stahelin, R. V. Membrane-protein interactions in cell signaling and membrane trafficking. *Annu. Rev. Biophys. Biomol. Struct.* **34,** 119–151 (2005).
14. Heo, W. D. *et al.* PI(3,4,5)P3 and PI(4,5)P2 lipids target proteins with polybasic clusters to the plasma membrane. *Science* **314,** 1458–1461 (2006).
15. Hilgemann, D. W. & Ball, R. Regulation of cardiac Na^+,Ca^{2+} exchange and K_{ATP} potassium channels by PIP2. *Science* **273,** 956–959 (1996).
16. Stanfield, P. R., Nakajima, S. & Nakajima, Y. Constitutively active and G-protein coupled inward rectifier K^+ channels: Kir2.0 and Kir3.0. *Rev. Physiol. Biochem. Pharmacol.* **145,** 47–179 (2002).
17. Rohács, T. *et al.* Specificity of activation by phosphoinositides determines lipid regulation of Kir channels. *Proc. Natl Acad. Sci. USA* **100,** 745–750 (2003).
18. Enkvetchakul, D., Jeliazkova, I. & Nichols, C. G. Direct modulation of Kir channel gating by membrane phosphatidylinositol 4,5-bisphosphate. *J. Biol. Chem.* **280,** 35785–35788 (2005).
19. Tao, X., Avalos, J. L., Chen, J. & MacKinnon, R. Crystal structure of the eukaryotic strong inward-rectifier K^+ channel Kir2.2 at 3.1 Å resolution. *Science* **326,** 1668–1674 (2009).
20. Zhang, H., He, C., Yan, X., Mirshahi, T. & Logothetis, D. E. Activation of inwardly rectifying K^+ channels by distinct PtdIns(4,5)P2 interactions. *Nature Cell Biol.* **1,** 183–188 (1999).
21. Cheng, W. W., D'Avanzo, N., Doyle, D. A. & Nichols, C. G. Dual-mode phospholipid regulation of human inward rectifying potassium channels. *Biophys. J.* **100,** 620–628 (2011).
22. Clarke, O. B. *et al.* Domain reorientation and rotation of an intracellular assembly regulate conduction in Kir potassium channels. *Cell* **141,** 1018–1029 (2010).
23. Pegan, S. *et al.* Cytoplasmic domain structures of Kir2.1 and Kir3.1 show sites for modulating gating and rectification. *Nature Neurosci.* **8,** 279–287 (2005).
24. Vagin, A. & Teplyakov, A. An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D* **56,** 1622–1624 (2000).
25. Collaborative Computational Projet 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
26. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
27. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
28. Xie, L.-H., John, S., a., Ribalet, B. & Weiss, J. N. Activation of inwardly rectifying potassium (Kir) channels by phosphatidylinosital-4,5-bisphosphate (PIP2): interaction with other regulatory ligands. *Prog. Biophys. Mol. Biol.* **94,** 320–335 (2007).

Author Contributions S.B.H. purified and crystallized the protein; collected, processed and refined crystallographic data, and performed electrophysiology experiments. X.T. aided in experimental design and provided assistance in all aspects of the project. R.M. designed the study and analysed data. All authors wrote and discussed the manuscript.

Author Information Atomic coordinates and structure factors for the reported crystal structures have been deposited into the Protein Data Bank under accession codes 3SPI (wild-type PIP2), 3SPC (wild-type PPA), 3SPH (PIP2(I223L)), 3SPJ (apo(I223L)) and 3SPG (PIP2(R186A)). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interest. Correspondence and requests for materials should be addressed to R.M. (mackinn@rockefeller.edu).

METHODS

Cloning, expression and purification. Kir2.2 from chicken with a GFP and a 1D4 epitope at the C terminus was expressed in *Pichia* and purified in n-decyl-β-D-maltopyranoside (DM, Anatrace) by 1D4 antibody affinity chromatography followed by PreScission protease cleavage and gel filtration on a superdex 200 column as previously described[19]. Purified protein was concentrated to 9 mg ml^{-1}. For crystallization trials of PIP$_2$–Kir2.2 channel complex, freshly prepared PIP$_2$ (10 mM stock in water) was added to the concentrated protein at a final concentration of 0.6–1 mM lipid and 8 mg ml^{-1} protein and incubated for about an hour before setting up trays. For the crystallization trials of the PPA–Kir2.2 channel complex, 5 mM PPA (100 mM stock in water) was used.

Structure determination. Co-crystals of Kir2.2 with PIP$_2$ or PPA were obtained from a 200 nl (100:100 nl protein:reservoir mixture) hanging drops. The protein buffer solution contained 4 mM DM, 20 mM dithiothreitol, 3 mM TCEP, 150 mM KCl and 20 mM Tris-HCl pH 8.0. Reservoir solution yielding the best diffracting crystals contained 0.3–0.6 M KCl, 50 mM HEPES (pH 6.5–7.5) plus 10–20% PEG 400 (w/v) or 3–8% PEG 4000 (w/v). Diamond-shaped crystals, 150–350 μm in the longest dimension, grew within 48 h at 4 °C. The crystals were cryoprotected in reservoir solution plus 25–30% (v/v) glycerol (5% increment steps) and flash frozen in liquid nitrogen. Diffraction data were collected at beamlines X29 and X25 (Brookhaven NSLS). Crystals with PIP$_2$ or PPA diffracted to 2.6–3.3 Å or 2.45 Å respectively. The crystals all belong to the *I*4 space group with one subunit in the asymmetric unit. Phases were obtained by molecular replacement from apo-Kir2.2 (PDB code 3JYC) using MolRep[24] in the CCP4 suite[25]. The models were built in Coot[26] and refined in Phenix[27] to an R_{free} of 0.22 to 0.28. There are no Ramachandran outliers (97% most favoured, 3% allowed). Complete crystallographic data and refinement statistics are shown in Supplementary Table 1. The PIP$_2$-bound model contains residues from 42–369. In the PPA-bound

structure, part of the interfacial helix is disordered and the final model contains residues 42–62 and 70–369. Waters were added with ARP/wARP[29] in the CCP4 suite[25] and manually adjusted in the 2.45 and 2.6 Å structures. Figures were made with PyMOL[30].

Electrophysiology. cRNA of chicken Kir2.2 was made from NdeI linearized Kir2.2 in the pGEM vector[19] using the Amplicap T7 RNA kit (Epicentre Biotechnologies). *Xenopus* oocytes were prepared as described[19] and injected with 50 nl of cRNA 12–20 h later. All recordings were made with patch clamp in inside-out configuration 2–3 days after injection. Injected oocytes were treated with ND96 (96 mM NaCl, 2 mM KCl, 1.8 mM CaCl$_2$, 1 mM MgCl$_2$, 50 μg ml^{-1} gentamycin, pH 7.6 with NaOH) plus 200 mM NaCl for 5–10 min and the vitelline membrane was removed before seal formation. On-cell membrane seals were formed using pipettes with typical resistance of 0.4–0.9 MΩ and large inside-out patches were excised with currents ranging from 0.2 to 5 nA and seals from 0.4 to 1 GΩ. The bath solution contained 130 mM KCl, 5 mM HEPES, 5 mM K$_2$EDTA, pH 7.4 with KOH. The pipette solution contained 140 mM KCl, 5 mM HEPES, 0.3 mM CaCl$_2$, 1 mM MgCl$_2$, pH 7.4 with KOH. For PIP$_2$ rescuing experiments described in Fig. 1b, 10 mM dioctanoyl PIP$_2$ prepared in water was added to the bath solution and mixed by pipetting.

All patch recordings were made with a voltage ramp from +80 to −80 mV in 10 s duration under the control of an Axopatch 200B amplifier, Digidata 1440A analogue-to-digital converter and pClamp10.1 software (Axon Instruments). For Fig. 1b, the voltage ramp was repeated every 30 s after patch excision and the amount of current at +70 mV was plotted against the time (immediately after excision: time = 0). Figure 1a and b was made with Igor Pro (Wavemetrics).

29. Cohen, S. X. *et al.* Towards complete validated models in the next generation of ARP/wARP. *Acta Crystallogr. D* **60**, 2222–2229 (2004).
30. Delano, W. L. PyMOL. ⟨http://www.pymol.org⟩ (Delano Scientific, 2002).

LETTER

In vitro centromere and kinetochore assembly on defined chromatin templates

Annika Guse[1], Christopher W. Carroll[1], Ben Moree[1], Colin J. Fuller[1] & Aaron F. Straight[1]

During cell division, chromosomes are segregated to nascent daughter cells by attaching to the microtubules of the mitotic spindle through the kinetochore. Kinetochores are assembled on a specialized chromatin domain called the centromere, which is characterized by the replacement of nucleosomal histone H3 with the histone H3 variant centromere protein A (CENP-A). CENP-A is essential for centromere and kinetochore formation in all eukaryotes but it is unknown how CENP-A chromatin directs centromere and kinetochore assembly[1]. Here we generate synthetic CENP-A chromatin that recapitulates essential steps of centromere and kinetochore assembly *in vitro*. We show that reconstituted CENP-A chromatin when added to cell-free extracts is sufficient for the assembly of centromere and kinetochore proteins, microtubule binding and stabilization, and mitotic checkpoint function. Using chromatin assembled from histone H3/CENP-A chimaeras, we demonstrate that the conserved carboxy terminus of CENP-A is necessary and sufficient for centromere and kinetochore protein recruitment and function but that the CENP-A targeting domain—required for new CENP-A histone assembly[2]—is not. These data show that two of the primary requirements for accurate chromosome segregation, the assembly of the kinetochore and the propagation of CENP-A chromatin, are specified by different elements in the CENP-A histone. Our unique cell-free system enables complete control and manipulation of the chromatin substrate and thus presents a powerful tool to study centromere and kinetochore assembly.

Metazoan centromeres are specified epigenetically by the presence of CENP-A nucleosomes[3]. Structural differences between CENP-A and histone H3 nucleosomes[2,4] and/or specific protein recognition elements in CENP-A seem to provide the information that specifies centromere identity and directs kinetochore assembly in a DNA-sequence-independent manner[5–10]. Moreover, many metazoan centromeres are complex in their organization, with interspersed blocks of CENP-A nucleosomes and histone H3 nucleosomes assembled on long arrays of repetitive DNA[11–13]. The difficulty in purifying and manipulating complex centromeres has limited our understanding of how centromeric chromatin promotes centromere and kinetochore formation and chromosome segregation.

To mimic the arrays of CENP-A nucleosomes present in complex vertebrate centromeres, we reconstituted human CENP-A chromatin from recombinant components (Fig. 1a). We generated saturated chromatin arrays by salt dialysis of purified histone proteins H2A, H2B, H4 and either CENP-A or H3 with a biotinylated DNA template containing 19 repeats of a 147 bp high-affinity nucleosome positioning sequence (19X601) (Supplementary Fig. 1a, b)[14,15]. We bound the biotinylated arrays to streptavidin-coated magnetic beads, thereby immobilizing the arrays so that they can be easily added to and recovered from cell extracts (Fig. 1a and Supplementary Fig. 1c–e).

We recently demonstrated that the essential centromere protein CENP-C directly recognizes the C terminus of CENP-A in mononucleosomes but not in isolated CENP-A$_2$/H4$_2$ tetramers[5] (our unpublished observations). Therefore, we tested *in vitro* translated human and *Xenopus laevis* CENP-C for binding to reconstituted H3 and CENP-A

chromatin. Human and *Xenopus* CENP-A are >50% identical (Supplementary Fig. 2a) and we find that both human and *Xenopus* CENP-C bind specifically to human CENP-A chromatin arrays *in vitro*, when compared to H3 chromatin arrays (Supplementary Fig. 2b).
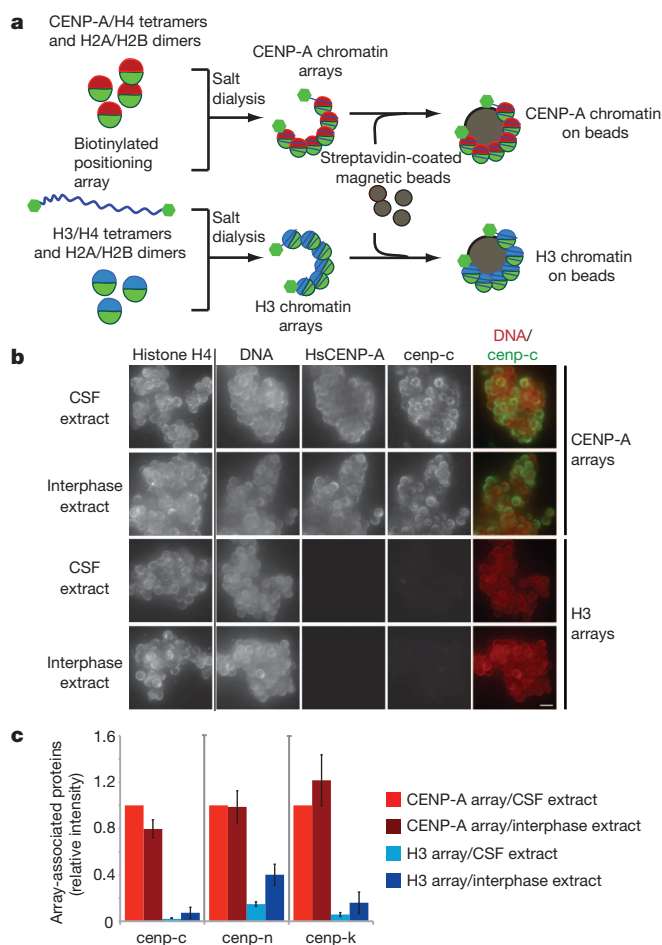
Figure 1 | Reconstituted CENP-A chromatin supports centromere assembly in *Xenopus* egg extracts. a, A schematic showing the reconstitution of CENP-A and H3 chromatin arrays and the attachment of the chromatin to magnetic beads via biotin end-labelled DNA. **b**, Representative images comparing cenp-c binding to human CENP-A (HsCENP-A) and H3 chromatin arrays in CSF and interphase *Xenopus* extract. The left column shows the separate histone H4 staining used for normalization of the quantification, followed by staining for DNA, human CENP-A and cenp-c. A merge image of the DNA (red) and cenp-c (green) channels is shown in the right column. Scale bar, 5 μm. **c**, Quantification of the array-associated centromeric proteins cenp-c, cenp-n and cenp-k in CSF and interphase extracts, normalized to histone H4 levels. The levels are rescaled so that CENP-A arrays in CSF are set at 1. Error bars represent the standard error of the mean (s.e.m.), $n = 3$ ($P < 0.05$ between CENP-A and H3 chromatin arrays for cenp-c, cenp-n and cenp-k).

[1]Department of Biochemistry, Stanford Medical School, Beckman 409A, Stanford, California 94305-5307, USA.

Xenopus egg extract is a widely used cell-free system to study chromosome segregation[16]. Egg extracts are arrested in metaphase II of meiosis by the activity of cytostatic factor (CSF) and the cell-cycle state of the extract can be transitioned into interphase by adding calcium. We developed a quantitative immunofluorescence assay to determine whether centromere proteins bound to CENP-A chromatin arrays when arrays were added to *Xenopus* egg extracts. CENP-N and CENP-K are centromere proteins that are required for proper centromere and kinetochore assembly in somatic cells, and we have previously shown that CENP-N, similar to CENP-C, directly binds to the CENP-A nucleosome[6]. We found that cenp-c, cenp-n and cenp-k specifically associated with CENP-A arrays independent of the cell-cycle stage of the extract (Fig. 1b, c and Supplementary Fig. 2c–f). The centromere protein cenp-t that binds to either H3 nucleosomes or DNA at centromeres did not selectively bind CENP-A chromatin arrays (Supplementary Fig. 3a, b)[17]. Similarly, the inner centromere protein incenp and polo-like kinase 1 (plk1) associated with both types of chromatin arrays (Supplementary Fig. 3c). *Xenopus* incenp is targeted to chromatin through phosphorylation of both H2A and H3 and thus may have affinity for both CENP-A and H3 chromatin[18–20] and plk1 associates with chromatin in *Xenopus* egg extract independent of the kinetochore[21]. Furthermore, reconstituted chromatin segments are unlikely to generate paired sister chromatids with inner centromeres because naked DNA and linear DNA replicates inefficiently in these egg extracts[22]. The specific recruitment of the centromere proteins cenp-c, cenp-n and cenp-k, however, indicates that reconstituted CENP-A chromatin arrays can support essential steps in the centromere assembly process *in vitro*.

Functional kinetochores assemble on sperm chromatin in metaphase *Xenopus* egg extract. At high sperm concentration, microtubule depolymerization causes mitotic checkpoint activation, resulting in the increased association of checkpoint proteins with kinetochores and cell-cycle arrest[23]. We tested whether reconstituted CENP-A chromatin arrays support kinetochore assembly and checkpoint protein binding after microtubule depolymerization. We added CENP-A or H3 arrays to CSF-arrested egg extracts and then cycled the extracts through interphase and back into mitosis, in the presence or absence of nocodazole, as outlined in Fig. 2a and demonstrated in Supplementary Fig. 4a. The constitutive centromere protein cenp-c and the microtubule-binding kinetochore protein ndc80 bound to CENP-A arrays in the presence or absence of nocodazole (Fig. 2b, c and Supplementary Fig. 4b). The spindle assembly checkpoint proteins cenp-e, mad2, rod (also known as kntc1) and zw10 associated with CENP-A chromatin at intermediate levels in the absence of nocodazole but upon microtubule depolymerization their binding increased 2–4 fold (Fig. 2b). Western blot analysis showed that cenp-c and ndc80 are precipitated with CENP-A arrays independent of microtubule depolymerization. *Xenopus* zw10 and rod are enriched on CENP-A arrays upon nocodazole treatment in metaphase, regardless of whether the extract has been cycled through interphase (Fig. 2c). These results indicate that CENP-A chromatin arrays respond to microtubule depolymerization by recruiting mitotic checkpoint proteins (Fig. 2b, c and Supplementary Fig. 4b).

Microtubule binding is a hallmark of kinetochore function and decondensed sperm chromatin efficiently supports spindle formation in egg extracts (Fig. 3a, left)[24]. However, chromatin assembled on naked DNA induces spindle formation in *Xenopus* egg extracts independent of kinetochores[25]. When we added CENP-A and H3 chromatin beads into mitotic egg extract we observed microtubule polymerization around the majority of CENP-A arrays but only around a subset of H3 arrays (Fig. 3a, left). We quantified the amount of microtubule polymer associated with each type of array and found significantly more microtubules associated with CENP-A chromatin beads (Fig. 3b and Supplementary Fig. 5a). This indicates that CENP-A chromatin preferentially stabilizes microtubules or promotes their polymerization. We observed heterogeneous microtubule structures around the CENP-A chromatin beads ranging from bipolar spindles to stabilized
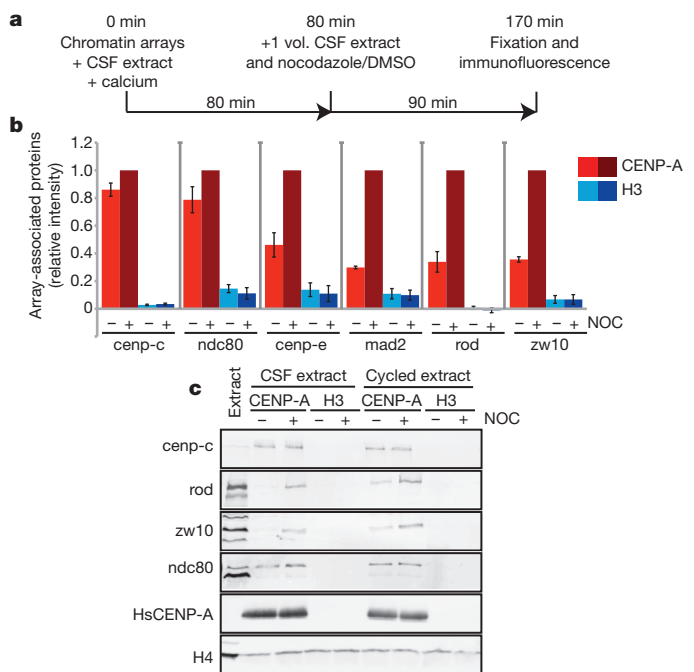
Figure 2 | CENP-A chromatin specifically recruits kinetochore proteins as a response to a mimic of kinetochore detachment from microtubules. a, A schematic showing the experimental procedure. **b**, Quantification of immunofluorescence analysis of cenp-c, ndc80, cenp-e, mad2, rod or zw10 recruitment to chromatin arrays with (+) and without (−) nocodazole (NOC). The levels are rescaled so that CENP-A arrays with nocodazole are set at 1. Error bars represent s.e.m., $n = 3$ ($P < 0.05$ between (−) and (+) nocodazole for cenp-e, mad2, rod and zw10 binding to CENP-A chromatin arrays). **c**, Western blot analysis of cenp-c, ndc80, rod and zw10 recruitment to CENP-A (HsCENP-A) and H3 chromatin arrays with and without nocodazole in CSF and cycled egg extracts. H4 levels are shown as a loading control.

microtubules or microtubule bundles (Fig. 3a and Supplementary Fig. 5a, b). A second property of functional kinetochores is that kinetochore-associated microtubule bundles (k-fibres) are stable to cold treatment, which depolymerizes non-kinetochore microtubules. We asked whether kinetochores assembled on CENP-A chromatin could stabilize microtubules to cold shock by incubating the microtubule assembly reactions for 10 min at 4 °C. We found that kinetochores assembled on CENP-A chromatin arrays stabilized microtubules to cold shock similar to kinetochores assembled on native sperm chromatin whereas H3 chromatin arrays did not (Fig. 3a, c and Supplementary Fig. 5c). When we completely depolymerized microtubules with nocodazole we observed mad2 recruitment to native sperm centromeres and CENP-A chromatin beads but not H3 chromatin beads (Fig. 3a, c and Supplementary Fig. 5c). These results indicate that CENP-A chromatin arrays, similar to native sperm chromatin, assemble functional kinetochores that promote microtubule binding, k-fibre stabilization and spindle checkpoint function (Fig. 3a).

In cells, unattached kinetochores activate the mitotic checkpoint and delay mitotic exit until all chromosomes are properly attached and aligned[26,27]. We tested whether kinetochores assembled on CENP-A chromatin arrays could generate a mitotic checkpoint response to microtubule depolymerization and delay the cell cycle. We mixed CENP-A and H3 chromatin with CSF extracts, cycled the reactions through interphase and then cycled them back into mitosis in the presence or absence of nocodazole (Fig. 2a). We then released the extract from mitosis into interphase a second time and monitored the kinetics of this transition by measuring the mitosis-specific phosphorylation of wee1 (phospho-wee1) (Fig. 3d). On release from mitosis, phospho-wee1 levels rapidly declined and were undetectable after 30 min in control extracts containing CENP-A chromatin or H3 chromatin, as well as in extracts containing H3 chromatin in the presence of
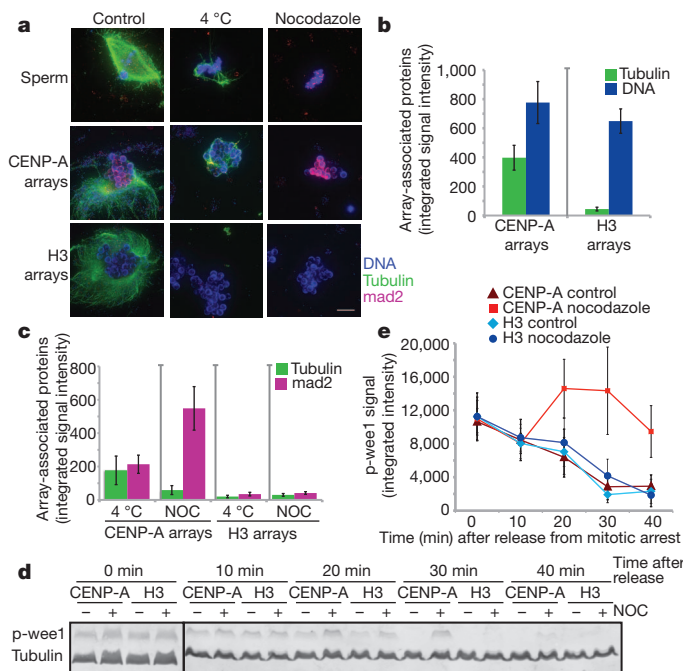
Figure 3 | Kinetochores assembled on reconstituted CENP-A chromatin bind microtubules and generate a mitotic checkpoint signal.
a, Representative images of microtubule polymerization induced by sperm or reconstituted CENP-A and H3 chromatin. Microtubules (green) and mad2 (magenta) levels are shown. Scale bar, 10 μm. b, Quantification of tubulin and DNA associated with CENP-A and H3 chromatin beads. Error bars represent s.e.m., $n = 5$. c, Quantification of tubulin and mad2 levels associated with CENP-A and H3 chromatin beads after cold shock (4 °C) and nocodazole (NOC) treatment. Error bars represent s.e.m., $n = 5$. d, Western blot showing phospho-wee1 (p-wee1) levels as an indicator of the cell-cycle stage and tubulin levels as a loading control. Samples from different time points after release from mitotic arrest are shown for CENP-A and H3 chromatin arrays, each incubated with nocodazole (+) or with DMSO (−) as a control. e, Quantification of four independent experiments showing the phospho-wee1 signal intensity (p-wee1 signal) over time (min). Error bars represent s.e.m., $n = 4$.

nocodazole (Fig. 3d, e). In extracts containing CENP-A chromatin and nocodazole, the phospho-wee1 signal increased until 20 min after calcium addition and subsequently declined until 40 min after calcium addition to a level only slightly lower than that before release (Fig. 3d, e). In the presence of CENP-A chromatin and nocodazole, cyclin B levels rapidly declined but then stabilized, similar to the response observed for native sperm chromatin[23]. However, cyclin B was not stabilized in the presence of H3 chromatin and nocodazole (Supplementary Fig. 5d, e). We estimate that the number of CENP-A nucleosomes we are adding to the egg extract exceeds the CENP-A nucleosome concentration required to activate the checkpoint using sperm nuclei[23]. The lower efficiency of reconstituted arrays for checkpoint signalling may be due to the comparatively short length of our reconstituted CENP-A chromatin to native CENP-A chromatin or the lack of replicated sister chromatids and inner centromeres important for tension-dependent checkpoint activation. Despite these differences, our synthetic CENP-A chromatin supports a mitotic checkpoint response that mimics the response of native kinetochores to microtubule depolymerization.

The reconstituted chromatin system we have developed provides a distinct experimental advantage over native metazoan centromeric chromatin because the chromatin template can be easily manipulated to dissect the roles of histone proteins in centromere function. A central question in centromere function is how CENP-A chromatin directs the assembly of the centromere and kinetochore. CENP-N recognizes the CATD region of the CENP-A nucleosome while CENP-C binds the C-terminal tail of CENP-A[5,6]. However, the relative importance of these two recognition mechanisms in centromere and kinetochore assembly is incompletely understood.

We generated chromatin arrays containing chimaeric CENP-A/H3 proteins to ask how the CENP-A CATD domain and the CENP-A C terminus influence centromere and kinetochore assembly (Fig. 4a). We characterized the level of histone exchange and/or loss from the arrays during incubation in extracts and found that the majority of recombinant human CENP-A nucleosomes were stable during the incubation, indicating low exchange and/or loss rates (Supplementary Fig. 6a, b). We detected a low level of phosphorylated histone H3 on CENP-A chromatin arrays in CSF extract (11.7% ± 7% compared to H3 arrays) and in extract that had been cycled through interphase and back into mitosis (22% ± 13% compared to H3 arrays) (Supplementary Fig. 6c, d). The chimaeric arrays containing CENP-A with the histone H3 tail (CENP-A + H3C) exhibited similar levels of exchange (Supplementary Fig. 6c, d). The Xenopus cenp-a present in the extract did not appreciably exchange onto any of the arrays (detection limit ~5–10% exchange) (Supplementary Fig. 6c). The absence of gross rearrangements or bulk histone exchange suggests that chromatin arrays can be used to dissect how individual domains of CENP-A influence kinetochore assembly.
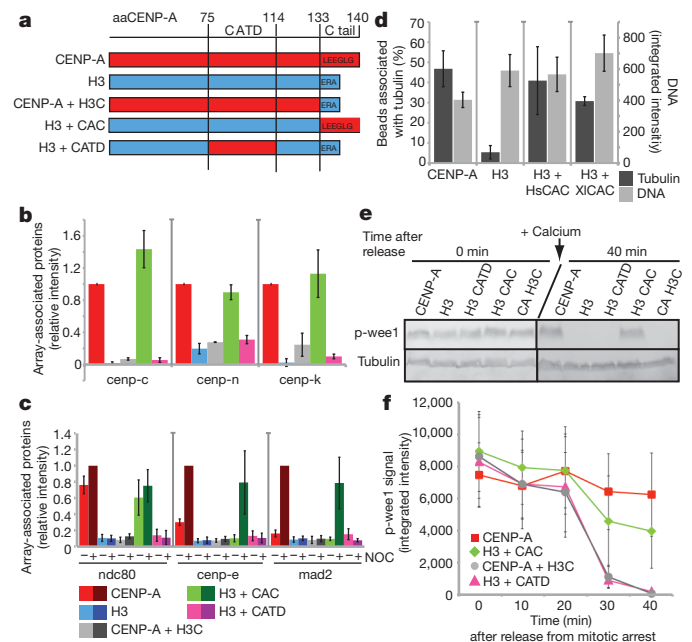
Figure 4 | The CENP-A C terminus is required for centromere and kinetochore assembly in Xenopus egg extract. a, A schematic showing the different CENP-A/H3 chimaeras used in this study. The numbers at the top represent the amino acid (aa) within human CENP-A. b, Quantification of immunofluorescence analysis of cenp-c, cenp-k and cenp-n recruitment to wild-type and chimaeric arrays. The relative amounts of each centromere protein bound to the arrays are shown relative to CENP-A arrays set to 1. Error bars represent s.e.m., $n = 3$ ($P \leq 0.05$ for all proteins binding to CENP-A arrays compared to chimaeric arrays except for the H3 arrays containing the CENP-A C terminus). c, Quantification of immunofluorescence analysis of ndc80, cenp-e, mad2 recruitment to chimaeric chromatin arrays with (+) and without (−) nocodazole (NOC). Values are displayed relative to CENP-A arrays in the presence of nocodazole set to 1. Error bars represent s.e.m., $n = 4$. The efficiencies of recruitment of kinetochore proteins to CENP-A and H3 + CAC arrays in nocodazole were not statistically distinguishable ($P \geq 0.26$ for ndc80, cenp-e and mad2). d, Quantification of microtubule binding to CENP-A, H3, H3 + human CAC (HsCAC) and H3 + Xenopus CAC (XlCAC) chromatin arrays represented as percentage of beads associated with tubulin levels above threshold (dark grey bars, left y-axis). Average DNA levels on chromatin beads are shown representing the levels of chromatin arrays bound to beads (light grey bars, right y-axis). Error bars represent s.e.m., $n = 4$ for CENP-A and H3 arrays, $n = 5$ for H3 + human CAC arrays and $n = 2$ for H3 + Xenopus CAC arrays. e, Western blot analysis shows phospho-wee1 (p-wee1) levels as an indicator of the cell-cycle stage at 0 min and 40 min after mitotic exit. Tubulin levels are shown as a loading control. f, Quantification of the phospho-wee1 signal intensity over time. Error bars represent s.e.m., $n = 5$.

Using our *in vitro* centromere and kinetochore assembly assay, we found that cenp-c bound with equal efficiency to chromatin arrays assembled with either wild-type CENP-A or with chimaeras of histone H3 with the CENP-A C-terminal six amino acids (H3 + CAC) but not CENP-A + H3C (Fig. 4b and Supplementary Fig. 7a, left). This demonstrates that the CENP-A C terminus is necessary and sufficient for recruiting cenp-c to CENP-A chromatin arrays in egg extracts, as it is for CENP-A mononucleosome binding *in vitro*[5].

Xenopus cenp-k depends on cenp-c for its association with sperm centromeres[28] and cenp-k also associated with the wild-type and H3 + CAC arrays (Fig. 4b and Supplementary Fig. 7a). Surprisingly, we found that H3 + CAC arrays recruited cenp-n as efficiently as wild-type CENP-A arrays, even though these arrays lack the CATD recognition element for CENP-N[6]. *Xenopus* cenp-n binding to either CENP-A + H3C or H3 + CATD arrays was no better than its binding to H3 chromatin arrays, indicating that the CENP-A C terminus is required for cenp-n association with CENP-A chromatin in *Xenopus* egg extract (Fig. 4b and Supplementary Fig. 7a). The lack of *Xenopus* cenp-n binding to H3 + CATD and CENP-A + H3C chromatin arrays is not due to species differences because *Xenopus* cenp-n binds human CENP-A mononucleosomes *in vitro* in the absence of CENP-C (Supplementary Fig. 7b). The association of cenp-n and cenp-k with chromatin arrays was dependent on cenp-c, as cenp-c depletion from the extract (Supplementary Fig. 8a) reduced the binding to background levels (Supplementary Fig. 8b, c). This was not due to depletion of cenp-n or cenp-k by cenp-c, as we have previously shown that complementation of cenp-c-depleted extracts restores cenp-k binding and CENP-K is known to depend on CENP-N for its centromere localization[6,28,29,30]. The dependence of CENP-N on CENP-C for its localization to CENP-A arrays may reflect a role for CENP-C in altering the geometry of centromeric chromatin to promote access of CENP-N to CENP-A nucleosomes, or it may reflect the assembly of CENP-N into the larger CCAN complex recruited to the centromere via CENP-C. Our results demonstrate that cenp-c recognition of the CENP-A C terminus is necessary and sufficient for cenp-n and cenp-k association with chromatin arrays in *Xenopus* egg extract.

We analysed the chromatin requirements for mitotic kinetochore formation using the experimental strategy illustrated in Fig. 2a. The kinetochore proteins ndc80, cenp-e, mad2, rod and zw10 are efficiently recruited to wild-type and H3 + CAC chromatin arrays, but not to CENP-A + H3C or H3 + CATD chromatin arrays (Fig. 4c, Supplementary Fig. 9a and Supplementary Fig. 10a). Similar to wild-type CENP-A chromatin, only the checkpoint proteins cenp-e, mad2, zw10 and rod increased in their association with H3 + CAC after microtubule depolymerization (Fig. 4c, Supplementary Fig. 9a and Supplementary Fig. 10a). As with wild-type CENP-A arrays, the H3 + CAC arrays showed increased associated microtubule polymer indicating that the C terminus of CENP-A directs the formation of microtubule binding or stabilization activity (Fig. 4d). Human and *Xenopus* CENP-A differ by two amino acids in their C-terminal tail (Supplementary Fig. 2a) and chimaeric nucleosome arrays containing the *Xenopus* C-terminal tail of cenp-a fused to H3 (H3 + *Xenopus* CAC) were equally efficient in cenp-c recruitment and microtubule binding as human H3 + CAC arrays (Fig. 4d and Supplementary Fig. 10b); indicating that the mode of interaction between CENP-C and CENP-A is conserved.

We assayed the ability of chimaeric nucleosome arrays to promote mitotic checkpoint arrest after microtubule depolymerization and found that H3 + CATD and CENP-A + H3C did not delay the exit from mitosis but that H3 + CAC did (Fig. 4e, f). The delay of mitotic exit caused by H3 + CAC arrays was less effective than that of CENP-A chromatin arrays, indicating that regions of CENP-A in addition to the C terminus increase the effectiveness of checkpoint signalling, possibly by stabilizing CCAN and kinetochore protein interactions with chromatin (Fig. 4e, f). Taken together, our data demonstrate that the primary chromatin determinant for functional centromere and kinetochore assembly is the C terminus of CENP-A and its recognition by CENP-C.

Here we have shown that reconstituted CENP-A chromatin, in the absence of native centromeric DNA, is necessary and sufficient for centromere and kinetochore assembly. Our data imply that short domains of CENP-A chromatin are sufficient for assembling core components of the centromere and kinetochore in the absence of higher-order organization of centromeric chromatin and interspersed domains of H3 chromatin.

Using our *in vitro* system, we have directly assessed how domains of CENP-A participate in centromere and kinetochore assembly, even when the mutations we analyse would be expected to be lethal *in vivo*. We find that the CENP-A C terminus is both necessary and sufficient for the recruitment of centromere and kinetochore proteins, for microtubule binding and for a checkpoint response to microtubule depolymerization. We suggest that CENP-A performs two functions that can be separated molecularly: (1) the CENP-A CATD provides a recognition mechanism for targeting of CENP-A to centromeres to maintain centromeric chromatin[2,6–8]; and (2) the CENP-A C-terminal tail domain recruits the conserved centromere protein CENP-C to promote centromere and kinetochore assembly[5]. We envision the use of more complex chromatin templates to understand the importance of higher-order chromatin organization and regulatory modifications in centromere assembly and function.

METHODS SUMMARY

Histone proteins and chimaeras were purified as described previously[5,6,15] and assembled onto a biotin end-labelled tandem array of 19 high-affinity nucleosome positioning sequences (19X601) by salt dialysis[14]. Chromatin arrays were bound to streptavidin-coated magnetic Dynabeads (Invitrogen). *X. laevis* extracts were prepared as previously described[16] and centromere protein binding to chromatin arrays was performed in freshly prepared CSF egg extract for 1 h with or without calcium addition. Arrays were fixed in formaldehyde and stained for centromere proteins by indirect immunofluorescence. Kinetochore and checkpoint protein assembly was assayed by adding arrays to extracts released into interphase with calcium for 80 min followed by re-addition of CSF extract in the presence or absence of nocodazole (10 μg ml^{-1}) for another 90 min. To analyse microtubule binding, chromatin arrays were incubated in CSF for 90 min. Reactions were sedimented through a glycerol cushion onto a coverslip followed by tubulin immunofluorescence. Chromatin-array-dependent inhibition of mitotic exit was assayed as described for kinetochore protein binding, but calcium was added a second time to release extracts into interphase. The cell-cycle state was monitored by western blotting using anti-phospho-wee1 antibody, provided by J. E. Ferrell.

Images were collected as 13 axial planes at 2 μm intervals on a Nikon Eclipse-80i microscope using a ×60, 1.4 NA PlanApo oil lens and a CoolSnapHQ CCD camera (Photometrics) with MetaMorph software (MDS Analytical Technologies). Axial stacks were maximum intensity projected and quantified using custom software. For normalization of each experiment, a separate histone H4 staining was performed to quantify the exact array coupling efficiency.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Cheeseman, I. M. & Desai, A. Molecular architecture of the kinetochore-microtubule interface. *Nature Rev. Mol. Cell Biol.* **9,** 33–46 (2008).
2. Black, B. E. *et al.* Structural determinants for generating centromeric chromatin. *Nature* **430,** 578–582 (2004).
3. Black, B. E. & Bassett, E. A. The histone variant CENP-A and centromere specification. *Curr. Opin. Cell Biol.* **20,** 91–100 (2008).
4. Sekulic, N., Bassett, E. A., Rogers, D. J. & Black, B. E. The structure of (CENP-A-H4)$_2$ reveals physical features that mark centromeres. *Nature* **467,** 347–351 (2010).
5. Carroll, C. W., Milks, K. J. & Straight, A. F. Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J. Cell Biol.* **189,** 1143–1155 (2010).
6. Carroll, C. W., Silva, M. C. C., Godek, K. M., Jansen, L. E. T. & Straight, A. F. Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. *Nature Cell Biol.* **11,** 896–902 (2009).
7. Dunleavy, E. M. *et al.* HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell* **137,** 485–497 (2009).
8. Foltz, D. R. *et al.* Centromere-specific assembly of CENP-A nucleosomes is mediated by HJURP. *Cell* **137,** 472–484 (2009).

9. Hu, H. *et al.* Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. *Genes Dev.* **25,** 901–906 (2011).
10. Tachiwana, H. *et al.* Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature.* doi:10.1038/nature10258 (10 July, 2011).
11. Blower, M. D., Sullivan, B. A. & Karpen, G. H. Conserved organization of centromeric chromatin in flies and humans. *Dev. Cell* **2,** 319–330 (2002).
12. Ribeiro, S. *et al.* A super-resolution map of the vertebrate kinetochore. *Proc. Natl Acad. Sci. USA* **107,** 10484–10489 (2010).
13. Zinkowski, R. P., Meyne, J. & Brinkley, B. R. The centromere-kinetochore complex: a repeat subunit model. *J. Cell Biol.* **113,** 1091–1110 (1991).
14. Huynh, V., Robinson, P. & Rhodes, D. A method for the *in vitro* reconstitution of a defined ''30 nm'' chromatin fibre containing stoichiometric amounts of the linker histone. *J. Mol. Biol.* **345,** 957–968 (2005).
15. Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M. & Richmond, T. J. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* **272,** 301–311 (1997).
16. Desai, A., Murray, A., Mitchison, T. J. & Walczak, C. E. The use of *Xenopus* egg extracts to study mitotic spindle assembly and function *in vitro. Methods Cell Biol.* **61,** 385–412 (1999).
17. Hori, T. *et al.* CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell* **135,** 1039–1052 (2008).
18. Kawashima, S. A., Yamagishi, Y., Honda, T., Ishiguro, K.-i. & Watanabe, Y. Phosphorylation of H2A by Bub1 prevents chromosomal instability through localizing shugoshin. *Science* **327,** 172–177 (2010).
19. Kelly, A. E. *et al.* Survivin reads phosphorylated histone H3 threonine 3 to activate the mitotic kinase Aurora B. *Science* **330,** 235–239 (2010).
20. Wang, F. *et al.* Histone H3 Thr-3 phosphorylation by Haspin positions Aurora B at centromeres in mitosis. *Science* **330,** 231–235 (2010).
21. Budde, P. P., Kumagai, A., Dunphy, W. G. & Heald, R. Regulation of Op18 during spindle assembly in *Xenopus* egg extracts. *J. Cell Biol.* **153,** 149–158 (2001).
22. Blow, J. J. & Laskey, R. A. Initiation of DNA replication in nuclei and purified DNA by a cell-free extract of *Xenopus* eggs. *Cell* **47,** 577–587 (1986).
23. Minshull, J., Sun, H., Tonks, N. K. & Murray, A. W. A MAP kinase-dependent spindle assembly checkpoint in *Xenopus* egg extracts. *Cell* **79,** 475–486 (1994).
24. Sawin, K. E. & Mitchison, T. J. Mitotic spindle assembly by two different pathways *in vitro. J. Cell Biol.* **112,** 925–940 (1991).
25. Heald, R. *et al.* Self-organization of microtubules into bipolar spindles around artificial chromosomes in *Xenopus* egg extracts. *Nature* **382,** 420–425 (1996).
26. Nicklas, R. B., Ward, S. C. & Gorbsky, G. J. Kinetochore chemistry is sensitive to tension and may link mitotic forces to a cell cycle checkpoint. *J. Cell Biol.* **130,** 929–939 (1995).
27. Rieder, C. L., Cole, R. W., Khodjakov, A. & Sluder, G. The checkpoint delaying anaphase in response to chromosome monoorientation is mediated by an inhibitory signal produced by unattached kinetochores. *J. Cell Biol.* **130,** 941–948 (1995).
28. Milks, K. J., Moree, B. & Straight, A. F. Dissection of CENP-C-directed centromere and kinetochore assembly. *Mol. Biol. Cell* **20,** 4246–4255 (2009).
29. Foltz, D. R. *et al.* The human CENP-A centromeric nucleosome-associated complex. *Nature Cell Biol.* **8,** 458–469 (2006).
30. McClelland, S. E. *et al.* The CENP-A NAC/CAD kinetochore complex controls chromosome congression and spindle bipolarity. *EMBO J.* **26,** 5033–5047 (2007).

METHODS

Histone expression. CENP-A/H4 and H3/H4 wild-type and chimaeric tetramers, as well as H2A and H2B dimers were expressed and purified as described previously[5,6,15,31].

Preparation of biotinylated array DNA. A tandem array of 19 copies of the high-affinity nucleosome positioning sequence (19X601)[14,32] was digested with EcoRI, XbaI, DraI and HaeII (NEB) overnight to excise the 19-nucleosome positioning sequence array and to digest the remaining backbone DNA to smaller DNA fragments. The array DNA was then purified by PEG precipitation and dialysed against 10 mM Tris-HCl pH 8.0, 0.25 mM EDTA as previously described[14].

The array DNA was end labelled with biotin by end filling the EcoRI and XbaI sites using Klenow DNA polymerase for 4 h at 37 °C in a reaction containing 35 μM Biotin-14-dATP (Invitrogen), α-thio-dTTP and α-thio-dGTP (Chemcyte) and dCTP. The labelled DNA was then purified using a PCR fragment purification kit (Qiagen). The biotinylation efficiency was determined by adding FITC-streptavidin (final concentration of 10 μg ml^{-1}) to 500 ng of purified array DNA and monitoring the fraction of gel-shifted DNA after migration in a 0.7% agarose gel.

Chromatin array assembly. To assemble chromatin arrays, biotinylated DNA, CENP-A/H4 or H3/H4 tetramers and H2A/H2B dimers were mixed at a stochiometry of 1:1.2:2 or 1:0.9:2.2, respectively, in high-salt buffer (10 mM Tris-HCl pH 7.5, 0.25 mM EDTA, 2 M NaCl) and then dialysed into low-salt buffer (10 mM Tris-HCl pH 7.5, 0.25 mM EDTA, 2.5 mM NaCl) over 60–70 h at 4 °C. Final array DNA concentration typically was 0.15 mg ml^{-1} to 0.2 mg ml^{-1}.

To assess the efficiency of nucleosome assemblies, arrays were digested at room temperature (approximately 22 °C) overnight with AvaI in a low-magnesium buffer (50 mM potassium acetate, 20 mM Tris-acetate, 0.5 mM magnesium acetate, 1 mM dithiothreitol, pH 7.9). Digested chromatin arrays were supplemented with glycerol (20% final concentration) and separated on a native 5% acrylamide gel in 0.5× Tris/Borate/EDTA buffer for 80 min at 10 mA. Gels were stained with EtBr (1 μg ml^{-1}) to visualize DNA.

Coupling of biotinylated chromatin arrays to Dynabeads. Biotinylated chromatin arrays were coupled to prewashed streptavidin-coated magnetic Dynabeads (Invitrogen) at a ratio of 10 μg DNA to 1 mg beads in 50 mM Tris-HCl pH 8.0, 75 mM NaCl, 0.25 mM EDTA, 2.5% polyvinyl alcohol (PVA) and 0.05% Triton-X-100 for 1–2 h. The beads were then equilibrated in 75 mM Tris-HCl pH 8.0, 75 mM NaCl, 0.25 mM EDTA, 0.05% Triton-X-100 and either used directly or stored at 4 °C for later use.

X. laevis egg extracts. X. laevis CSF extracts were prepared as previously described[16,33]. To assess the binding of centromeric proteins to chromatin arrays in CSF and interphase egg extracts, chromatin arrays were mixed with freshly prepared CSF egg extract with or without CaCl$_2$ (final concentration 0.6 mM) at a nucleosome concentration of ~100 nM unless stated otherwise. The reactions were incubated for 1 h at 4 °C or at 16–20 °C in a water bath, the arrays were re-isolated from extracts by exposure to a magnet and then washed three times in 1× CSF-XB buffer (10 mM HEPES pH 7.7, 2 mM MgCl$_2$, 0.1 mM CaCl$_2$, 100 mM KCl, 5 mM EGTA, 50 mM sucrose) supplemented with 0.05% Triton-X-100. Chromatin arrays were fixed in CSF-XB buffer, 0.05% Triton-X-100, 2% formaldehyde for 5 min. After fixation, chromatin arrays were washed into antibody dilution buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Triton-X-100, 2% BSA) and analysed by immunofluorescence.

Kinetochore and spindle checkpoint protein assembly were analysed by mixing chromatin arrays with CSF extract and CaCl$_2$ (final concentration 0.6 mM). Reactions were incubated at 16–20 °C for 80 min to allow extracts to release into interphase and mixed every 15 min. One volume of fresh CSF extract was added together with nocodazole (or DMSO) at 10 μg ml^{-1} and samples were held at 16–20 °C for another 90 min. After 170 min total incubation time, samples for immunofluorescence analysis were washed and fixed as described above.

The cell-cycle state was verified by loading 2 μl extract of all relevant time points onto SDS–PAGE, followed by western blotting using the anti-phospho-wee1 antibody[34].

To assess the ability of chromatin arrays to inhibit mitotic exit, arrays were mixed with CSF extract and CaCl$_2$ (final concentration: 0.6 mM). The samples were incubated for 80 min to induce the release into interphase. In the next step, one volume of fresh CSF extract, supplemented with nocodazole/DMSO, was added to cycle the extract back into a mitotic arrest. After 90 min, CaCl$_2$ was added again to release the extract from mitotic arrest. Western blot samples were taken at all indicated time points and processed as described.

To analyse microtubule binding by CENP-A and H3 chromatin arrays, chromatin arrays were mixed with CSF extract and incubated for 90 min at 18–20 °C. During incubation samples were mixed every 15 min. Reactions were fixed for 10 min in 2.5% formaldehyde, sedimented through a glycerol cushion onto coverslips and post-fixed for 5 min in ice-cold methanol followed by immunofluorescence analysis[35]. To assay for mad2 levels and microtubule stabilization, reactions were either supplemented with nocodazole at a final concentration of 10 μg ml^{-1} or shifted to 4 °C for 10 min after the 90 min incubation time.

Immunodepletion. Depletion of Xenopus cenp-c from Xenopus egg extracts was performed as described previously[28].

Cloning and antibody generation. The X. laevis cenp-n cDNA clone (GenBank accession number BC084956) was purchased from American Type Culture Collection. Peptides against Xenopus cenp-n (acetyl-CPHKARNSFKITEKR-amide) were synthesized by Bio-Synthesis and peptide antibodies were generated as previously described[36].

Immunofluorescence. For immunofluorescence analysis, fixed chromatin arrays were bound to poly-L-lysine-coated acid-washed coverslips. The following primary antibodies were used for immunofluorescence staining and typically incubated at 4 °C overnight: anti-human CENP-A[30] was directly coupled to Alexa 647 (Molecular Probes), anti-H4 (Abcam), anti-Xenopus cenp-c, anti-Xenopus cenp-e, anti-Xenopus cenp-k and anti-Xenopus cenp-n and anti-tubulin (Dm1α; Sigma). Rabbit antibodies were generated against the full-length Xenopus polo kinase made in Sf9 cells and a GST fusion to the first 379 amino acids of Xenopus incenp made in E. coli. The anti-mad2 antibody was provided by A. Murray (Harvard University), and R.-H. Chen (Institute of Molecular Biology, Academia Sinica), the anti-Xenopus zw10 and anti-Xenopus rod antibodies were provided by G. Kops (University Medical Center Utrecht) and the anti-Xenopus ndc80 antibody was provided by P. Todd Stukenberg (University of Virginia). Alexa-conjugated secondary antibodies were used at 1 μg ml^{-1} (Molecular Probes). Propidium iodide at 1 μg ml^{-1} or Hoechst at 10 μg ml^{-1} was used to visualize DNA.

Microscopy and analysis. Images were collected on a Nikon Eclipse 80i microscope using a ×60, 1.4 NA Plan Apo VC oil immersion lens, a Sedat Quad filter set (Chroma Technology) using MetaMorph software (MDS Analytical Technologies) and a charge-coupled device camera (CoolSnapHQ; Photometrics). Thirteen axial planes at 2 μm intervals were acquired with an MFC-2000 Z-axis drive (Applied Scientific Instrumentation). Axial stacks were maximum intensity projected and then quantified using custom software (Matlab) to identify beads in each image and to quantify the integrated intensity for each channel after background subtraction. Briefly, the propidium iodide stained (DNA) channel was used to find beads. Bead centroids were found by filtering the image using a structuring element that had a peak at a 17 pixel radial distance from the structuring element centre, corresponding to the bright ring seen around the edges of the beads. A 35 pixel diameter circle around the centroid of each bead identified was used as the region of interest for that bead. After beads were identified, regions of interest were transferred automatically to the remaining channels and the integrated signal intensity was calculated for each bead in each channel, normalized to the area of the bead region (which was uniform except in cases of partially overlapping beads), and background corrected using an average of three bead-sized regions manually chosen to be away from any beads. For each experiment, at least three images per coverslip were acquired and 20–300 beads were analysed per image. For the normalization of each experiment, a separate histone H4 staining was performed to quantify the exact coupling efficiency for each type of chromatin array and for each experiment.

Immunofluorescence microscopy images of the microtubule binding assays that were subjected to deconvolution were acquired with an Olympus IX70 microscope. The microscope was outfitted with a Deltavision Core system (Applied Precision) using an Olympus ×60 1.4NA Plan Apo lens, a Sedat Quad filter set (Semrock) and a CoolSnap HQ CCD Camera (Photometrics). The microscope was controlled via softworx 4.1.0 software (Applied Precision) and images were deconvolved using softworx v. 4.1.0 (Applied Precision). Microtubule quantification was performed using a modification of the same software used for centromere protein quantification.

Immunoblotting. Western blot samples were separated by SDS–PAGE and transferred onto PVDF membrane (Bio-Rad) in CAPS transfer buffer (10 mM 3-(cyclohexylamino)-1-propanesulfonic acid, pH 11.3, 0.1% SDS and 20% methanol). The following primary antibodies were typically incubated overnight at 4 °C: anti-Xenopus cenp-c[28], anti-tubulin (Dm1α, Sigma), anti-H4 (Abcam), anti-phospho H3 (Ser10) (Millipore), anti-phospho-wee1. The anti-phospho-wee1 antibody was provided by J. E. Ferrell (Stanford University)[34]. For additional primary antibodies, western blot samples were transferred onto PVDF membrane (Bio-Rad) in 20 mM Tris-Base, 200 mM glycine. Alexa fluorophore conjugated anti-rabbit or anti-mouse secondary antibodies (Molecular Probes) were used according to manufacturer's specification. Fluorescence was detected on a Typhoon 9400 Variable Mode Imager (Amersham Biosciences) and quantified using ImageJ (http://rsb.info.nih.gov/ij/). Actin antibodies were provided by J. Theriot (Stanford University) and anti-cyclin B was purchased from Santa Cruz Biotechnology.

In vitro binding of centromere proteins to chromatin arrays. Human and Xenopus CENP-C were in vitro translated (IVT) in rabbit reticulocyte extracts in the presence of 10 mCi ml^{-1} [^{35}S]methionine (Perkin Elmer) using the TnT Quick-Coupled Transcription/Translation system (Promega) according to the manufacturer's instructions. For a binding reaction (60 μl total volume), 5 μl of each IVT protein were mixed with chromatin arrays in bead buffer (75 mM Tris-HCl pH 7.5, 50 mM NaCl, 0.25 mM EDTA, 0.05% Triton-X-100). The final nucleosome concentration per reaction was 60 nM. Reactions were incubated at

4 °C for 1 h. The beads were washed three times with bead buffer and resuspended in 4× SDS loading buffer. Samples were separated on a SDS–PAGE, Coomassie stained and after drying scanned using a phosphorimager (Typhoon 4200, Amersham Biosciences) and quantified using ImageJ (http://rsb.info.nih.gov/ij/).

Statistical analysis. In each experiment, the relative levels of proteins associated with the chromatin arrays were normalized to values for wild type CENP-A arrays set to 1. For calculation of P values each data set was anchored at 1 and then log transformed followed by calculation of P values using a Student's t-test[37].

31. Luger, K., Rechsteiner, T. & Richmond, T. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* **304,** 3–19 (1999).

32. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276,** 19–42 (1998).

33. Murray, A. W. Cell cycle extracts. *Methods Cell Biol.* **36,** 581–605 (1991).

34. Kim, S., Song, E., Lee, K. & Ferrell, J. Jr. Multisite M-phase phosphorylation of *Xenopus* Wee1A. *Mol. Cell. Biol.* **25,** 10580–10590 (2005).

35. Hannak, E. & Heald, R. Investigating mitotic spindle assembly and function *in vitro* using *Xenopus laevis* egg extracts. *Nature Protocols* **1,** 2305–2314 (2006).

36. Field, C. M., Oegema, K., Zheng, Y., Mitchison, T. J. & Walczak, C. E. Purification of cytoskeletal proteins using peptide antibodies. *Methods Enzymol.* **298,** 525–541 (1998).

37. Osborne, J. W. *Best Practices in Quantitative Methods* (Sage Publications, 2008).

LETTER

Antibiotic resistance is ancient

Vanessa M. D'Costa[1,2]*, Christine E. King[3,4]*, Lindsay Kalan[1,2], Mariya Morar[1,2], Wilson W. L. Sung[4], Carsten Schwarz[3], Duane Froese[5], Grant Zazula[6], Fabrice Calmels[5], Regis Debruyne[7], G. Brian Golding[4], Hendrik N. Poinar[1,3,4] & Gerard D. Wright[1,2]

The discovery of antibiotics more than 70 years ago initiated a period of drug innovation and implementation in human and animal health and agriculture. These discoveries were tempered in all cases by the emergence of resistant microbes[1,2]. This history has been interpreted to mean that antibiotic resistance in pathogenic bacteria is a modern phenomenon; this view is reinforced by the fact that collections of microbes that predate the antibiotic era are highly susceptible to antibiotics[3]. Here we report targeted metagenomic analyses of rigorously authenticated ancient DNA from 30,000-year-old Beringian permafrost sediments and the identification of a highly diverse collection of genes encoding resistance to β-lactam, tetracycline and glycopeptide antibiotics. Structure and function studies on the complete vancomycin resistance element VanA confirmed its similarity to modern variants. These results show conclusively that antibiotic resistance is a natural phenomenon that predates the modern selective pressure of clinical antibiotic use.

Recent studies of modern environmental and human commensal microbial genomes have a much larger concentration of antibiotic resistance genes than has been previously recognized[4–6]. In addition, metagenomic studies have revealed diverse homologues of known resistance genes broadly distributed across environmental locales. This widespread dissemination of antibiotic resistance elements is inconsistent with a hypothesis of contemporary emergence and instead suggests a richer natural history of resistance[2]. Indeed, estimates of the origin of natural product antibiotics range from 2 Gyr to 40 Myr ago[7,8], suggesting that resistance should be similarly old. Previous publications claim to have cultured resistant bacteria from Siberian permafrost (for example ref. 9), but these results remain contentious (see Supplementary Information).

To determine whether contemporary resistance elements are modern or whether they originated before our use of antibiotics, we analysed DNA sequences recovered from Late Pleistocene permafrost sediments. The samples were collected east of Dawson City, Yukon, at the Bear Creek (BC) site (Fig. 1); prominent forms of ground ice (ice wedges and surface icings) are preserved in the exposure, immediately overlain by a distinctive volcanic ash layer, the Dawson tephra[10,11] (Supplementary Table 1 and Supplementary Figs 1 and 2). The tephra has been dated at several sites in the area to about 25,300 radiocarbon (^{14}C) years BP, or about 30,000 calendar years[10,12]. The cryostratigraphic context is similar to other sites in the area preserving relict permafrost and indicates that the permafrost has not thawed since the time of deposition (Supplementary Information). In the absence of fluid leaching, the site represents an ideal source of uncontaminated and securely dated ancient DNA.

Two frozen sediment cores (BC1 and BC4), 10 cm apart, were obtained 50 cm below the tephra. In accordance with appropriate protocols[13], we monitored contamination introduced during coring by spraying the drilling equipment and the outer surface of the cores

with high concentrations of *Escherichia coli* harbouring the *gfp* (green fluorescent protein) gene from *Aequorea victoria* (Supplementary Information).

After fracturing of the samples (Supplementary Fig. 3), total DNA was extracted from a series of five subsamples taken along the radius of each core (Supplementary Information). Quantitative polymerase
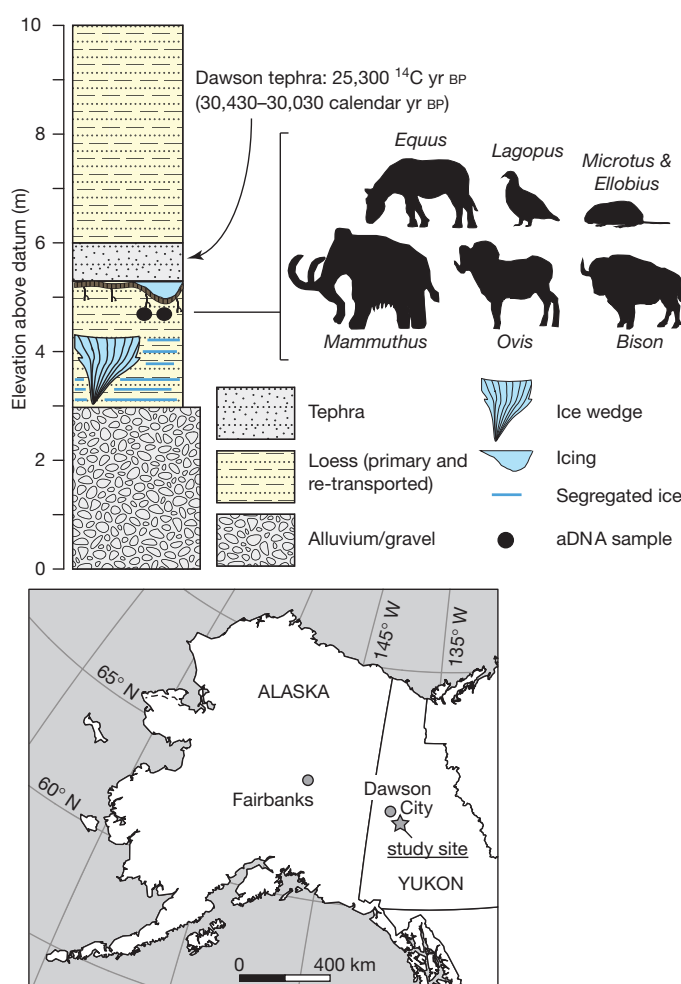
Figure 1 | Stratigraphic profile and location of Bear Creek site. Elevation is given in metres above base of exposure. Permafrost samples from below Dawson tephra were dated to about 30 kyr BP. Preservation of the ice below and above the sample indicates that the sediments have not thawed since deposition. Silhouettes represent mammals and birds identified from ancient DNA sequences that are typical of the regional Late Pleistocene environment. aDNA, ancient DNA.

[1]Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario, Canada, L8N 3Z5. [2]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, Canada, L8N 3Z5. [3]McMaster Ancient DNA Centre, Department of Anthropology, McMaster University, Hamilton, Ontario, Canada, L8S 4L9. [4]Department of Biology, McMaster University, Hamilton, Ontario, Canada, L8S 4K1. [5]Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Alberta, Canada, T6G 2E3. [6]Yukon Palaeontology Program, Department of Tourism and Culture, Yukon Government, PO Box 2703, Whitehorse, Yukon, Canada, Y1A 2C6. [7]Muséum National d'Histoire Naturelle, UMR 7206 Eco-anthropologie, 57 rue Cuvier, CP139, 75231 Paris cedex 05, France.
*These authors contributed equally to this work.

chain reaction (qPCR) analysis confirmed extremely high yields of *gfp* on both core exteriors, with 0.1% or less of this amount at the centre (Supplementary Information and Supplementary Fig. 4). This supports negligible leaching or cross-contamination during subsampling.

A crucial step lending support for the authenticity of the ancient DNA was to confirm the presence of DNA derived from flora and fauna characteristic of a late Pleistocene age, and the absence of common modern or Holocene floral and faunal sources. To explore the vertebrate and plant diversity, we amplified fragments of the mitochondrial 12S rRNA and chloroplast *trn*L and *rbc*L genes (Supplementary Table 3). Amplicons were sequenced with the 454 GS-FLX platform and identified by BLAST analysis of GenBank sequences (Supplementary Information).

The vertebrate sequences included abundant Late Pleistocene megafauna such as *Bison*, *Equus* and *Ovis*, as well as rodents (*Microtus* and *Ellobius*) and the rock ptarmigan, *Lagopus mutus* (Supplementary Fig. 6 and Supplementary Table 5). *Mammuthus* was detectable at low copy numbers with the use of a mammoth-specific qPCR assay, which is consistent with the low ratio of these fossils relative to bison and horse in the region[11,14]. The *rbc*L and *trn*L sequences revealed many plant groups that are also well documented in Beringia, including the grasses *Poa* and *Festuca*, sage (*Artemisia*) and willow (*Salix*)[15] (Supplementary Figs 7 and 8, and Supplementary Tables 6 and 7). No sequences of common Holocene vertebrates (for example elk or moose) or plants (for example spruce) were identified despite sequence conservation across the primer-binding sites; these results are consistent with other reports[16] that have argued against DNA leaching in permafrost sediments.

We focused our investigation of bacterial 16S rRNA sequences on the Actinobacteria, known for their ability to synthesize diverse secondary metabolites and for harbouring antibiotic resistance genes[4]. Deep sequencing of 16S amplicons (Supplementary Information) revealed genera commonly found in soil and permafrost microbial communities[17], including *Aeromicrobium*, *Arthrobacter* and *Frankia* (Supplementary Fig. 9 and Supplementary Table 8). Analysis of contaminant 16S sequences derived from extraction and PCR control reactions (Supplementary Table 4) suggested that these do not contribute to the ancient DNA data set; in fact not only were the copy numbers 1,000–30,000-fold lower than from the permafrost extracts, but with the exception of unclassified bacteria there was also very little overlap in the genera identified (Supplementary Fig. 9 and

Supplementary Table 8). Querying the permafrost sequences against the contaminant data set with the use of BLAST further confirmed their disparity: only 1% of the reads had 95–100% identity to a contaminant sequence, with a single sequence showing 100% identity.

We next developed a series of assays to detect genes encoding resistance to several major classes of antibiotic and representing diverse strategies of drug evasion (for example target modification, target protection and enzymatic drug inactivation) (Supplementary Information). Determinants included the ribosomal protection protein TetM, which confers resistance to tetracycline antibiotics by weakening the interaction between the drug and the ribosome; the D-Ala-D-Ala dipeptide hydrolase VanX, which is a component of the vancomycin resistance operon; the aminoglycoside-antibiotic-modifying acetyltransferase AAC(3); a penicillin-inactivating β-lactamase Bla (a member of the TEM group of β-lactamases); and the ribosome methyltransferase Erm, which blocks the binding of macrolide, lincosamide and type B streptogramin antibiotics. Amplification of *vanX*, *tetM* and *bla* fragments was successful, and triplicate PCR products from multiple extracts were cloned and multiple clones were sequenced.

The β-lactamase sequences demonstrated amino-acid identities between 53% and 84% with known determinants and clustered with one of two groups of enzymes: characterized β-lactamases from streptomycetes and uncharacterized β-lactamase-like hydrolytic proteins (Fig. 2a and Supplementary Fig. 14). We identified several *tetM*-related genes in the permafrost, most of which were most closely related to the actinomycete subset of ribosomal protection proteins, including the biochemically characterized self-resistance element OtrA from the oxytetracycline producer *Streptomyces rimosus*[18] (Fig. 2b). Most intriguing was the identification of *vanX* gene fragments, which spanned the entire phylogenetic space of characterized vancomycin resistance determinants found in the clinic and in the environment. These branch away from the cellular dipeptidases that are the likely progenitors the *vanX* family (Supplementary Fig. 10).

Vancomycin resistance took the clinical community by surprise when it emerged in pathogenic enterococci in the late 1980s[19]. In both clinical pathogens[19] and contemporary soil environments[4], resistance results from the acquisition of a three-gene operon *vanH–vanA–vanX* (*vanHAX*). These enzymes collectively reconstruct bacterial peptido-glycan to terminate in D-alanine-D-lactate in place of the canonical D-alanine-D-alanine, which is required for vancomycin binding and
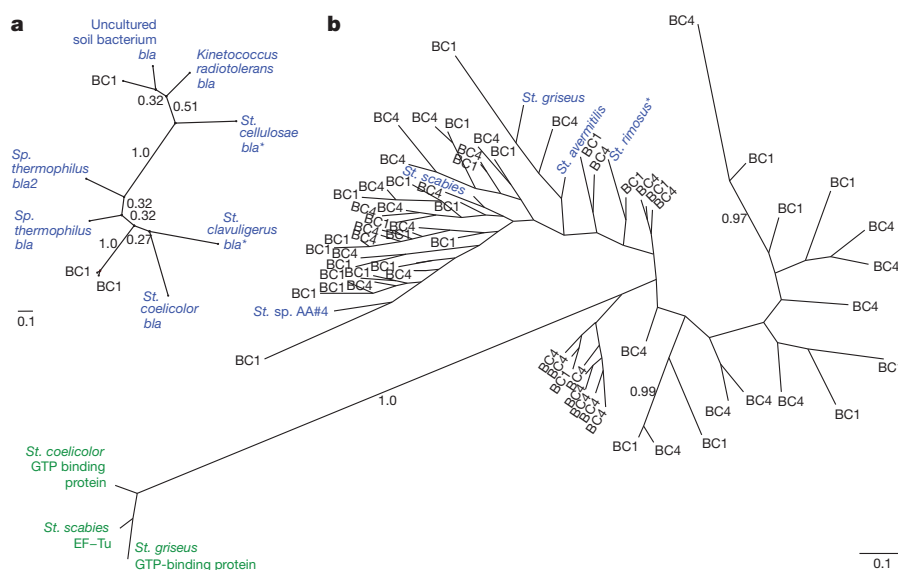
Figure 2 | Genetic diversity of ancient antibiotic resistance elements.
a, **b**, Unrooted Bayesian phylogenies of translated β-lactamase (*bla*) (**a**) and tetracycline resistance (*tetM*) (**b**). Blue denotes predicted resistance enzymes, and green those associated with other functions; permafrost-derived sequences are labelled with the originating core name. Sequences in which resistance activity has been biochemically verified are noted with a single asterisk (Supplementary Information). The scale bar represents 0.1 substitutions per site. Posterior probabilities are shown for **a**, and those of 0.7 or more are indicated for **b**. All unlabelled tips derive from ancient sequences. BC1, Bear Creek sample 1; BC4, Bear Creek sample 4.

subsequent antibiotic action. Although most forms of resistance are attributed to a single gene, this complex mechanism is exclusively associated with resistance and thus its presence provides unambiguous confirmation of its role as a resistance determinant.

With few exceptions, the *vanHAX* operon is invariant in genetic organization; it therefore offers a matchless template for confirming its presence with PCR assays that span the *vanHA* and *vanAX* boundaries. Two short qPCR assays were designed to confirm this contiguity
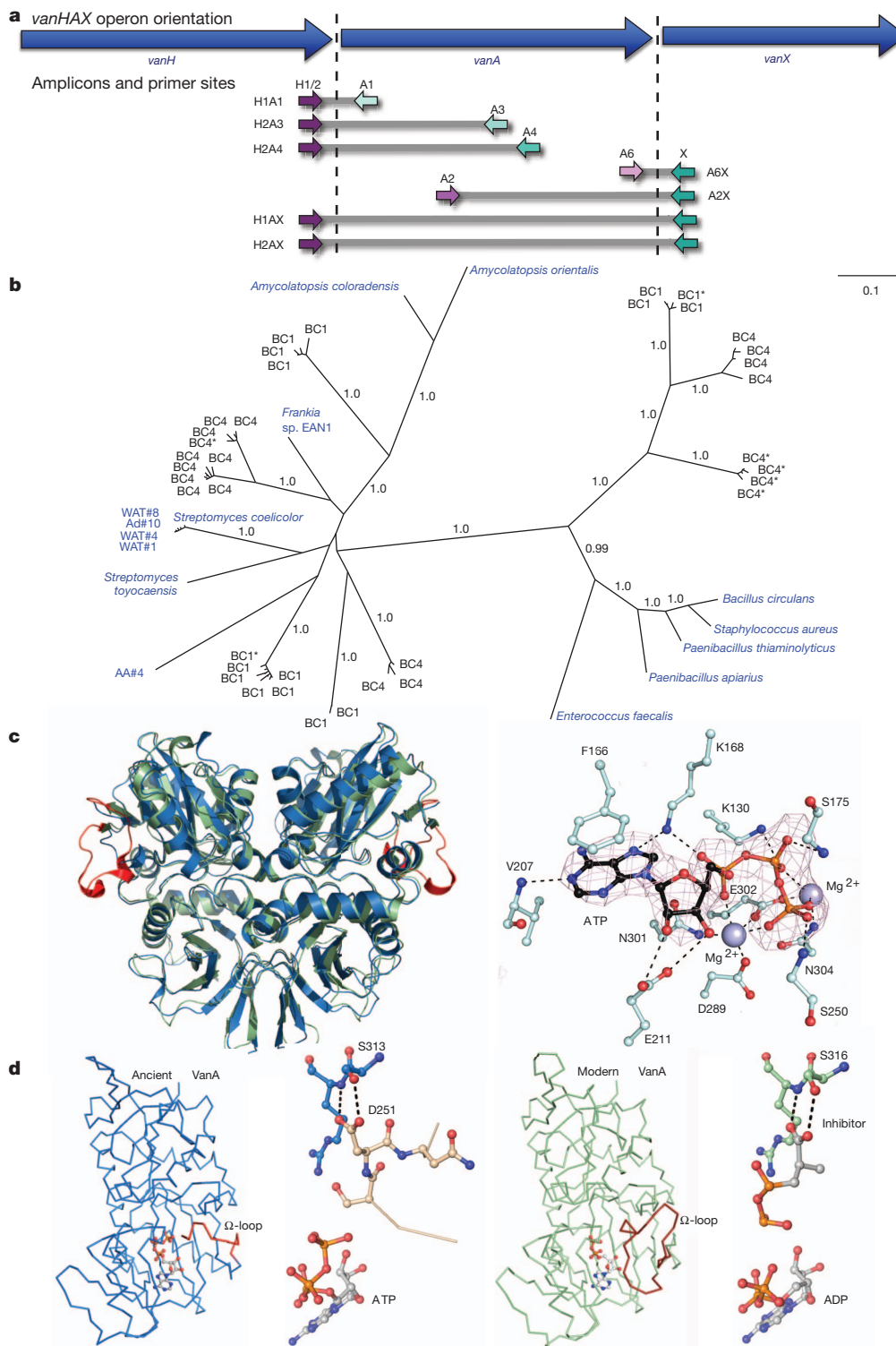
Figure 3 | Ancient vancomycin resistance elements. a, *vanHAX* amplicons used in this study, with primer names noted above each arrow. **b**, Unrooted Bayesian phylogeny of translated *vanA* sequences; blue denotes strains with *vanHAX* clusters confirmed to confer resistance; sequences containing stop codons but homology throughout are noted with a single asterisk (Supplementary Information). BC1, Bear Creek sample 1; BC4, Bear Creek sample 4. **c**, VanA$_{A2}$ structure. Left: ribbon diagram of the VanA$_{A2}$ dimer (blue)

overlaid with modern VanA (green), where the Ω-loop is coloured red; right: ball-and-stick representation of ATP binding. The electron density shown is an $F_o - F_c$ map contoured at 3σ. **d**, Comparison of modern and ancient VanA monomer structures. The Ω-loop is coloured red and detailed in the ball-and-stick figures. Ligands are shown in grey. Dashed lines represent hydrogen bonds.

Table 1 | *vanHAX* permutation tests

Amplicon	Number	Length (base pairs)	Probability of similarity by chance alone to *Streptomyces coelicolor* genes		
			vanH	*vanA*	*vanX*
H1A1	164	203–213	3.59×10^{-3}	4.39×10^{-17}	0.24
H1A1*	12	209–216	2.83×10^{-3}	8.16×10^{-16}	0.28
H2A3	24	573–605	9.83×10^{-3}	1.27×10^{-54}	0.22
H2A4	79	666–681	4.33×10^{-3}	6.15×10^{-53}	0.18
A6X	159	170–179	0.11	6.87×10^{-8}	5.64×10^{-9}
A6X*	11	176–179	0.04	2.96×10^{-8}	3.63×10^{-9}
A2X	96	735–796	0.11	1.80×10^{-59}	1.35×10^{-6}
HAX†	40	1,173–1,204	5.95×10^{-3}	9.32×10^{-92}	6.47×10^{-7}

* Clones from independent replication in France. †Includes both H1AX and H2AX.

(Fig. 3a and Supplementary Information). Positive results, including particularly high yields of the smallest amplicon, A6X (Supplementary Table 9), encouraged us to attempt amplification across both boundaries (that is, the complete *vanA* gene) in a single 1.2-kilobase amplicon. We also targeted fragments anchored on either boundary and extending as far as possible into *vanA*. None of the sequences from these products, or those generated by an independent laboratory (Supplementary Information), were present in GenBank. No contaminants were detected in more than 300 control reactions.

Phylogenetic analyses showed that many of the ancient *vanHAX* sequences cluster with characterized glycopeptide-resistant strains of Actinobacteria containing *vanHAX* cassettes (for example streptomycetes, glycopeptide-producing *Amycolatopsis* species and the nitrogen-fixing *Frankia* sp. EAN1pec) (Fig. 3b and Supplementary Figs 11 and 12). Another group falls between the actinobacterial sequences and the Firmicutes-derived cluster, which includes environmental *Paenibacillus* isolates and the pathogenic *Enterococci*, and may reflect an intermediate group.

Permutation tests were performed with the PRSS algorithm[20] (1,000 permutations each) to confirm that the sequences were statistically similar to those of vancomycin resistance genes (*vanHAX*) present in modern *Streptomyces*. As shown in Table 1, all *vanHA*-spanning clones have significant similarity to *vanH* and *vanA*, and all *vanAX*-spanning clones have significant similarity to *vanA* and *vanX*.

To ascertain whether the complete *vanA* sequences are indeed functional and do not represent PCR artefacts or pseudogenes, we synthesized four open reading frames from the 40 H1AX/H2AX sequences (Supplementary Information). Two of these generated soluble proteins suitable for purification to homogeneity. Enzymatic characterization indicated that these ligases were indeed D-alanine-D-lactate-specific (Supplementary Fig. 13), and analysis revealed steady-state kinetic parameters consistent with contemporary enzymes derived from both the clinic and the environment (Supplementary Table 10). These results clearly show that the *vanHAX* genes identified in the ancient samples encode enzymes capable of genuine antibiotic resistance.

We further confirmed the link between 30,000-year-old VanA and contemporary enzymes by determining the three-dimensional structure of VanA$_{A2}$ by X-ray crystallography (Supplementary Table 11 and Supplementary Information). The quaternary and tertiary structures of VanA$_{A2}$, crystallized in the ATP-bound form, show the overall D-Ala-D-X ligase fold of modern enzymes including VanA from vancomycin-resistant *Enterococcus faecium* (Fig. 3c, d). Superposition of ancient and modern VanA (Fig. 3c, d) reveals conservation of quaternary and tertiary structure with minor differences in Mg^{2+} and ATP γ-phosphate coordination. The Ω-loop comprises the biggest structural change; 13 amino-terminal residues (233–246) are absent from the electron density map of VanA$_{A2}$, including His 241 (His 244 in modern VanA), responsible for the lactate selectivity. The last seven Ω-loop residues (247–253) have clear electron density, undergoing a drastic 13 Å shift. These structural differences, however, are not reflected in enzyme function.

This work firmly establishes that antibiotic resistance genes predate our use of antibiotics and offers the first direct evidence that antibiotic resistance is an ancient, naturally occurring phenomenon widespread in the environment. This is consistent with the rapid emergence of resistance in the clinic and predicts that new antibiotics will select for pre-existing resistance determinants that have been circulating within the microbial pangenome for millennia. This reality must be a guiding principle in our stewardship of existing and new antibiotics.

METHODS SUMMARY

Permafrost cores were collected at Bear Creek, Yukon, then shipped frozen to the McMaster Ancient DNA Centre and stored at −40 °C. All subsequent procedures before PCR/qPCR amplification were performed in dedicated clean rooms, physically separated from laboratories containing modern DNA, bacterial cultures and amplification products. Contaminant leaching into the centre of cores after sampling was monitored by qPCR assays designed to detect *E. coli* DNA encoding the jellyfish green fluorescent protein sprayed onto coring equipment and the external surfaces of all collected cores. DNA was extracted from the centre of subsampled permafrost cores. PCR assays were designed to target vertebrates, plants, bacteria and specific antibiotic resistance elements. All products were sequenced with either the 454 GS-FLX platform or by standard cloning and sequencing procedures (GenBank accession numbers JN316287–JN366376). The ancient *vanA* gene identified from the permafrost was synthesized and expressed in *E. coli*, and the His$_6$-tagged protein was purified by immobilized metal-affinity chromatography. This protein was used in enzymatic studies to determine steady-state kinetics and was also studied by crystallography using the vapour-diffusion hanging-drop method. Data were collected at the National Synchrotron Light Source, Brookhaven National Laboratory, beamline X25 (PDB 1E4E).

1. Livermore, D. M. Has the era of untreatable infections arrived? *J. Antimicrob. Chemother.* **64**, i29–i36 (2009).
2. Wright, G. D. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature Rev. Microbiol.* **5**, 175–186 (2007).
3. Hughes, V. M. & Datta, N. Conjugative plasmids in bacteria of the 'pre-antibiotic' era. *Nature* **302**, 725–726 (1983).
4. D'Costa, V. M., McGrann, K. M., Hughes, D. W. & Wright, G. D. Sampling the antibiotic resistome. *Science* **311**, 374–377 (2006).
5. Dantas, G., Sommer, M. O. A., Oluwasegun, R. D. & Church, G. M. Bacteria subsisting on antibiotics. *Science* **320**, 100–103 (2008).
6. Sommer, M. O. A., Dantas, G. & Church, G. M. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128–1131 (2009).
7. Baltz, R. H. Antibiotic discovery from actinomycetes: will a renaissance follow the decline and fall? *SIM News* **55**, 186–196 (2005).
8. Hall, B. G. & Barlow, M. Evolution of the serine β-lactamases: past, present and future. *Drug Resist. Updat.* **7**, 111–123 (2004).
9. Mindlin, S. Z., Soina, V. S., Petrova, M. A. & Gorlenko, Z. M. Isolation of antibiotic resistance bacterial strains from Eastern Siberia permafrost sediments. *Russ. J. Genet.* **44**, 27–34 (2008).
10. Froese, D. G., Zazula, G. D. & Reyes, A. V. Seasonality of the late Pleistocene Dawson tephra and exceptional preservation of a buried riparian surface in central Yukon Territory, Canada. *Quat. Sci. Rev.* **25**, 1542–1551 (2006).
11. Froese, D. G. *et al.* The Klondike goldfields and Pleistocene environments of Beringia. *GSA Today* **19**, 4–10 (2009).
12. Brock, F., Froese, D. G. & Roberts, R. G. Low temperature (LT) combustion of sediments does not necessarily provide accurate radiocarbon ages for site chronology. *Quat. Geochronol.* **5**, 625–630 (2010).
13. Willerslev, E., Hansen, A. J. & Poinar, H. N. Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends Ecol. Evol.* **19**, 141–147 (2004).
14. Harington, C. R. & Clulow, F. V. Pleistocene mammals from Gold Run Creek, Yukon Territory. *Can. J. Earth Sci.* **10**, 697–759 (1973).
15. Zazula, G. D. *et al.* Ice-age steppe vegetation in east Beringia. *Nature* **423**, 603 (2003).

16. Haile, J. *et al.* Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proc. Natl Acad. Sci. USA* **106,** 22352–22357 (2009).
17. Gilichinsky, D. *et al.* in *Psychrophiles: From Biodiversity to Biotechnology* (eds Margesin, R., Schinner, F., Marx, J.-C. & Gerday, C.) 83–102 (Springer, 2008).
18. Doyle, D., McDowall, K. J., Butler, M. J. & Hunter, I. S. Characterization of an oxytetracycline-resistance gene, *otrA*, of *Streptomyces rimosus. Mol. Microbiol.* **5,** 2923–2933 (1991).
19. Courvalin, P. Vancomycin resistance in gram-positive cocci. *Clin. Infect. Dis.* **42,** S25–S34 (2006).
20. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85,** 2444–2448 (1988).

LETTER

The genome of the green anole lizard and a comparative analysis with birds and mammals

Jessica Alföldi[1]*, Federica Di Palma[1]*, Manfred Grabherr[1], Christina Williams[2], Lesheng Kong[3], Evan Mauceli[1], Pamela Russell[1], Craig B. Lowe[4], Richard E. Glor[5], Jacob D. Jaffe[1], David A. Ray[6], Stephane Boissinot[7], Andrew M. Shedlock[8], Christopher Botka[9], Todd A. Castoe[10], John K. Colbourne[11], Matthew K. Fujita[3,12], Ricardo Godinez Moreno[12], Boudewijn F. ten Hallers[13], David Haussler[14], Andreas Heger[3], David Heiman[1], Daniel E. Janes[12], Jeremy Johnson[1], Pieter J. de Jong[13], Maxim Y. Koriabine[13], Marcia Lara[1], Peter A. Novick[15], Chris L. Organ[12], Sally E. Peach[1], Steven Poe[16], David D. Pollock[10,17], Kevin de Queiroz[18], Thomas Sanger[12], Steve Searle[19], Jeremy D. Smith[6], Zachary Smith[11], Ross Swofford[1], Jason Turner-Maier[1], Juli Wade[20], Sarah Young[1], Amonida Zadissa[19], Scott V. Edwards[12], Travis C. Glenn[21], Christopher J. Schneider[22], Jonathan B. Losos[12], Eric S. Lander[1], Matthew Breen[2,23,24], Chris P. Ponting[3] & Kerstin Lindblad-Toh[1,25]

The evolution of the amniotic egg was one of the great evolutionary innovations in the history of life, freeing vertebrates from an obligatory connection to water and thus permitting the conquest of terrestrial environments[1]. Among amniotes, genome sequences are available for mammals and birds[2–4], but not for non-avian reptiles. Here we report the genome sequence of the North American green anole lizard, *Anolis carolinensis*. We find that *A. carolinensis* microchromosomes are highly syntenic with chicken microchromosomes, yet do not exhibit the high GC and low repeat content that are characteristic of avian microchromosomes[2]. Also, *A. carolinensis* mobile elements are very young and diverse—more so than in any other sequenced amniote genome. The GC content of this lizard genome is also unusual in its homogeneity, unlike the regionally variable GC content found in mammals and birds[5]. We describe and assign sequence to the previously unknown *A. carolinensis* X chromosome. Comparative gene analysis shows that amniote egg proteins have evolved significantly more rapidly than other proteins. An anole phylogeny resolves basal branches to illuminate the history of their repeated adaptive radiations.

The amniote lineage divided into the ancestral lineages of mammals and reptiles ~320 million years ago. Today, the surviving members of those lineages are mammals, comprising ~4,500 species, and reptiles, containing ~17,000 species. Within the reptiles, the two major clades diverged ~280 million years ago: the lepidosaurs, which contains lizards (including snakes) and the tuatara; and the archosaurs, containing crocodilians and birds (the position of turtles remains unclear)[6]. For simplicity, we will refer here to lepidosaurs as lizards (Fig. 1).

The study of the major genomic events that accompanied the transition to a fully terrestrial life cycle has been assisted by the sequencing of several mammal (K.L.-T. *et al.*, manuscript submitted) and three bird genomes[2–4]. The genome of the lizard *A. carolinensis* thus fills an important gap in the coverage of amniotes, splitting the long branch between mammals and birds and allowing more robust evolutionary analysis of amniote genomes.

For instance, almost all reptilian genomes contain microchromosomes, but these have only been studied at a sequence level in birds[2,7], raising the question as to whether the avian microchromosomes' peculiar sequence features are universal across reptilian microchromosomes[8]. Another example is the study of sex chromosome evolution. Nearly all placental and marsupial mammals share homologous sex chromosomes (XY)[9] and all birds share ZW sex chromosomes. However, lizards exhibit either genetic or temperature-dependent sex determination[10]. Characterization of lizard sex chromosomes would allow the study of previously unknown sex chromosomes and comparison of independent sex chromosome systems in closely related species.

Anolis lizards comprise a diverse clade of ~400 described species distributed throughout the Neotropics. These lizards have radiated, often convergently, into a variety of ecological niches with attendant morphological adaptations, providing one of the best examples of adaptive radiation. In particular, their diversification into multiple replicate niches on diverse Caribbean islands via interspecific competition and natural selection has been documented in detail[11]. *A. carolinensis* is the only anole native to the USA and can be found from Florida and Texas up to North Carolina. We chose this species for genome sequencing because it is widely used as a reptile model for experimental ecology, behaviour, physiology, endocrinology, epizootics and, increasingly, genomics.

The green anole genome was sequenced and assembled (AnoCar 2.0) using DNA from a female *A. carolinensis* lizard (Supplementary Tables 1–4). Fluorescence *in situ* hybridization (FISH) of 405 bacterial artificial chromosome (BAC) clones (from a male) allowed the assembly scaffolds to be anchored to chromosomes (Supplementary Table 5 and Supplementary Fig. 1). The *A. carolinensis* genome has been reported to have a karyotype of $n = 18$ chromosomes, comprising six pairs of large macrochromosomes and 12 pairs of small microchromosomes[12]. The draft genome sequence is 1.78 Gb in size (see Supplementary Table 3 for assembly statistics) and represents an intermediate

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [2]Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, North Carolina 27606, USA. [3]MRC Functional Genomics Unit, University of Oxford, Department of Physiology, Anatomy and Genetics, Oxford OX1 3QX, UK. [4]Stanford University School of Medicine, Department of Developmental Biology, Stanford, California 94305, USA. [5]University of Rochester, Rochester, New York 14607, USA. [6]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, Mississippi 39762, USA. [7]Department of Biology, Queens College, the City University of New York, New York, New York 11367, USA. [8]Biology Department and Graduate Program in Marine Genomics, College of Charleston, Charleston, South Carolina 29424, USA. [9]Harvard Medical School, Boston, Massachusetts 02115, USA. [10]Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, Colorado 80045, USA. [11]The Center for Genomics and Bioinformatics, Indiana University, Bloomington, Indiana 47405, USA. [12]Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA. [13]Children's Hospital Oakland Research Institute, Oakland, California 94609, USA. [14]Center for Biomolecular Science and Engineering, University of California, Santa Cruz, California 95064, USA. [15]Department of Biological Sciences and Geology, Queensborough Community College, Bayside, New York 11364, USA. [16]Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA. [17]Program in Computational Bioscience, University of Colorado School of Medicine, Aurora, Colorado 80045, USA. [18]Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, District of Columbia 20560, USA. [19]Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. [20]Departments of Psychology and Zoology, Program in Neuroscience, Michigan State University, East Lansing, Michigan 48824, USA. [21]Department of Environmental Health Science and Georgia Genomics Facility, University of Georgia, Athens, Georgia 30602, USA. [22]Biology Department, Boston University, Boston, Massachusetts 02215, USA. [23]Center for Comparative Medicine and Translational Research, North Carolina State University, Raleigh, North Carolina 27695, USA. [24]University of North Carolina Lineberger Comprehensive Cancer Center, Chapel Hill, North Carolina 27514, USA. [25]Science for Life Laboratory Uppsala, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala 751 23, Sweden.
*These authors contributed equally to this work.
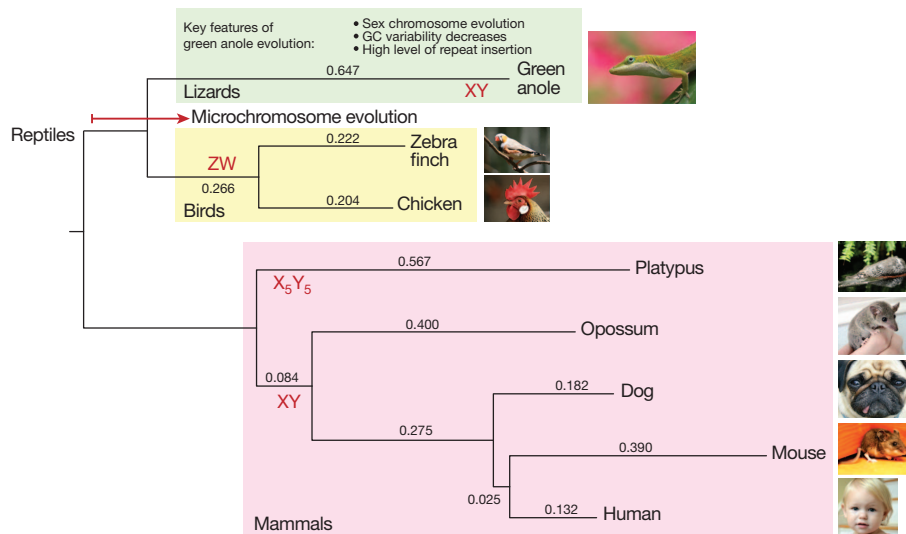
Figure 1 | Amniote phylogeny based on protein synonymous sites showing major features of amniote evolution. Major characteristics of lizard evolution including homogenization of GC content, high sex chromosome turnover and high levels of repeat insertion are featured. Sex chromosome inventions are indicated in red. Branch length is proportional to dS (the synonymous substitution rate); dS of each branch is indicated above the line.

between genome assemblies of birds (0.9–1.3 Gb) and mammals (2.0–3.6 Gb).

We find that few chromosomal rearrangements occurred in the 280 million years since anole and chicken diverged, as had been hinted at by previous comparisons using *Xenopus* and chicken[13]. There are 259 syntenic blocks (defined as consecutive syntenic anchors that are consistent in order, orientation and spacing, at a resolution of 1 Mb) between lizard and chicken (Supplementary Table 6 and Supplementary Fig. 2). Interestingly, 19 out of 22 anchored chicken chromosomes are each syntenic to a single *A. carolinensis* chromosome over their entire lengths (Fig. 2a); by contrast, only 6 (of 23) human chromosomes are syntenic to a single opossum chromosome over their entire lengths, even though the species diverged only 148 million years ago[14]. Segmental duplications follow trends seen in other amniote genomes (Supplementary Note, Supplementary Table 7 and Supplementary Fig. 3).

Approximately 30% of the *A. carolinensis* genome is composed of mobile elements, which comprise a much wider variety of active repeat families than is seen for either bird[2] or mammalian[15] genomes. The most active classes are long interspersed (LINE) elements (27%) and short interspersed (SINE) elements (16%)[16] (Supplementary Table 8). The majority of LINE repeats belong to five groups (L1, L2, CR1, RTE and R4) and seem to be recent insertions based on their sequence similarity (divergence ranges from 0.00–0.76%; ref. 17). This contrasts with observations of mammalian genomes, where only a single family of LINEs—L1—has predominated over tens of millions of years. The DNA transposons comprise at least 68 families belonging to five superfamilies: hAT, Chapaev, Maverick, Tc/Mariner and Helitron[18]. As with retrotransposons, the majority of DNA transposon families seem to be relatively young in contrast to the extremely few recently active DNA transposons found in other amniote genomes (Supplementary Table 9). Overall, *A. carolinensis* mobile elements feature significantly higher GC content (43.5%, $P < 10^{-20}$) than the genome-wide average of 40.3%. In addition to mobile elements, *A. carolinensis* exhibits a high density (3.5%) of tandem repeats, with length and frequency distributions similar to those of human microsatellite DNA[15]. We now know that amniote genomes come in at least three types: mammalian genomes are enriched for L1 elements and have a high degree of mobile element accumulation, bird genomes are repeat poor with very little mobile element activity, while the lizard genome contains an extremely wide diversity of active mobile element families but has a low rate of accumulation, which is reminiscent of the mobile element profile of teleostean fishes[19].

Most reptile genomes contain microchromosomes, but the numbers vary among species; the *A. carolinensis* genome contains 12 pairs of microchromosomes[12], whereas the chicken genome contains 28 pairs. Bird microchromosomes have very distinctive properties compared to bird macrochromosomes, such as higher GC and lower repeat contents[2], whereas lizard microchromosomes do not exhibit these features (Fig. 2b). Remarkably, all sequence anchored to microchromosomes in *A. carolinensis* also aligns to microchromosomes in the chicken genome, and all but one *A. carolinensis* microchromosome is syntenic to only a single corresponding chicken microchromosome (Fig. 2a). Microchromosomes conserved between *A. carolinensis* and chicken thus could have arisen in the reptile ancestor, whereas the remaining chicken microchromosomes could be derived in the bird lineage. Alternatively, the remaining chicken microchromosomes could have been present in the reptile ancestor but fused to form macrochromosomes in the lizard lineage.

The *A. carolinensis* genome has surprisingly little regional variation of GC content, substantially less than previously observed for birds and mammals; it is the only amniotic genome known whose nucleotide composition is as homogenous as the frog genome[5] (Supplementary Figs 4 and 5). Figure 3 illustrates how local GC content is evolutionarily conserved between human chromosome 14 and chicken chromosome 5, but to a much lesser degree with *A. carolinensis* chromosome 1. As all sequenced amniote genomes other than *A. carolinensis* contain these homologous varying levels of GC content ('isochores')[20], the ancestral amniote GC heterogeneity is likely to have eroded towards homogeneity in this lizard's lineage. It has been proposed that isochores with high GC content are a consequence of higher rates of GC-biased gene conversion in regions of higher recombination[2]. The greater GC homogeneity in the anole genome may thus reflect more uniform recombination rates, or else a substantially reduced bias towards GC during the resolution of gene conversion events in the *A. carolinensis* lineage (for a discussion, see ref. 5).

Both temperature-dependent sex determination and XY genetic sex determination have been found in Iguania[10]. Within the genus *Anolis*, there are species with heteromorphic XY chromosomes (including those with multiple X and Y chromosomes), and others with entirely homomorphic chromosomes[12]. *A. carolinensis* is known to have genetic sex determination[21], but the form of its sex chromosomes (ZW or XY) has thus far been unknown owing to a lack of obviously heteromorphic chromosomes.

In depth examination of male and female cells using FISH allowed us to identify the microchromosome previously designated as 'b' as the
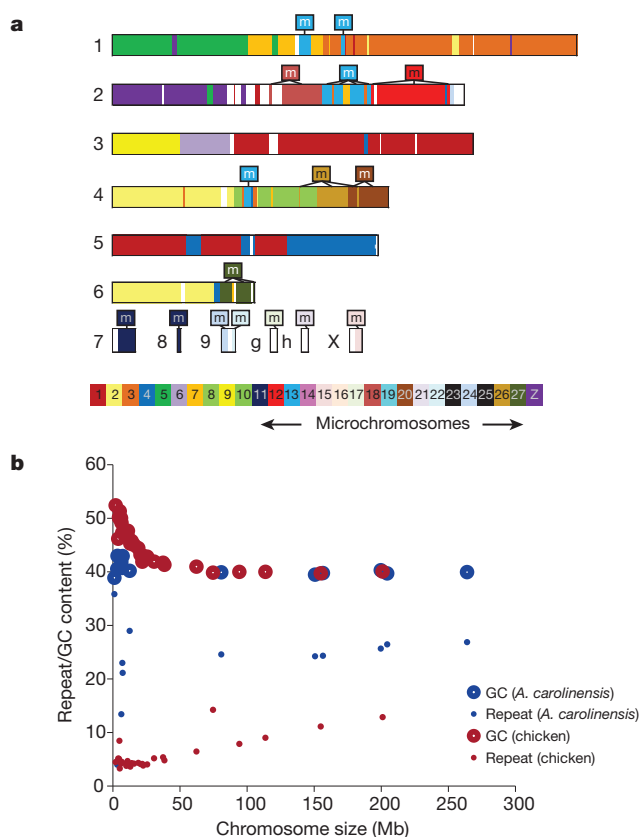
Figure 2 | A. carolinensis–chicken synteny map reveals synteny of reptile microchromosomes but dissimilar GC and repeat content. a, Very few rearrangements have occurred in the 280 million years since *A. carolinensis* and chicken diverged. *A. carolinensis* microchromosomes are exclusively syntenic to chicken microchromosomes. Horizontal coloured bars depict the six *A. carolinensis* macrochromosomes (1–6) and the six (of 12) *A. carolinensis* microchromosomes that have sequence anchored to them that is syntenic to the chicken genome (7, 8, 9, X, LGg, LGh). Chromosomes that could be ordered by size were assigned a number; the smaller microchromosomes that could not be distinguished by size were assigned a lowercase letter. Each colour corresponds to a different chicken chromosome as indicated in the key. Any part of an *A. carolinensis* chromosome that is syntenic to a chicken microchromosome is indicated by 'm'. **b**, Chicken microchromosomes have both higher GC content and lower repeat content than chicken macrochromosomes, whereas *A. carolinensis* chromosomes do not vary in GC or repeat content by chromosome size. Large circles designate the GC percentage of each chromosome in the chicken and lizard genomes with greater than 100 kb of sequence anchored to it. Small circles designate the percentage of the genome made up of repetitive sequence of each chromosome in the chicken (blue circles) and lizard (red circles) genomes.

A. carolinensis X chromosome; it is present in two copies in females and one in males. This chromosome is syntenic to chicken microchromosome 15. Eleven BACs assigned to two scaffolds, 154 (3.3 Mb) and chrUn0090 (1.8 Mb), hybridize via FISH to the p arms of the two X chromosomes in females, and hybridize to the p arm of the single X chromosome in males (Fig. 4 and Supplementary Fig. 1). *A. carolinensis* thereby shows a pattern representative of a male heterogametic system of genotypic sex determination. We have not identified the Y chromosome, but we hypothesize that *A. carolinensis* possesses both X and Y chromosomes, as both male and female cells contain the same number of chromosomes.

The 5.1 Mb of sequence assigned to the X chromosome contains 62 protein-coding genes (Supplementary Table 10); Gene Ontology (GO) terms associated with these genes show no significant enrichment. It is very likely that there is more X chromosome sequence that is currently labelled as unanchored scaffolds in the AnoCar 2.0 assembly. Identification of the *A. carolinensis* sex determination gene will require
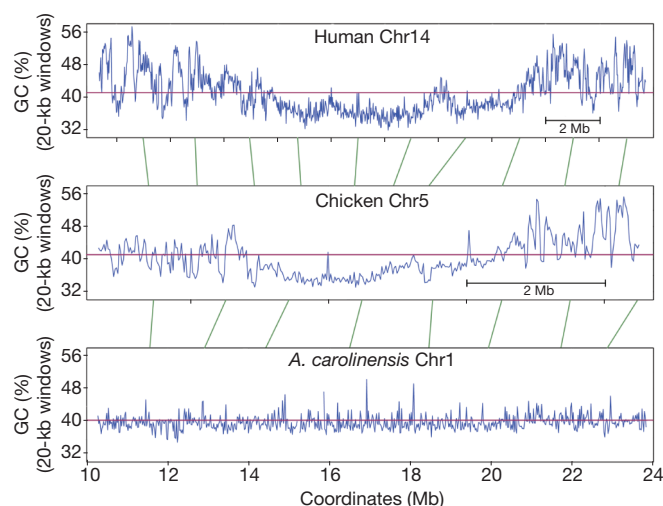
Figure 3 | The *A. carolinensis* genome lacks isochores. The *A. carolinensis* genome shows only very local variation in GC content, unlike the human and chicken genomes, which also show larger trends in GC variation, sometimes called isochores. Syntenic regions of human chromosome 14, chicken chromosome 5 and *A. carolinensis* chromosome 1 are shown. The human and chicken regions are inverted and rearranged to align with the *A. carolinensis* region. Blue lines depict GC percentage in 20-kb windows. The purple line designates the genome average. Green lines represent examples of syntenic anchors between the three genomes.

considerable functional biology, but we note that the chicken sex determination gene *DMRT1* is located on *A. carolinensis* chromosome 2 and that *SOX3* (the X chromosome paralogue of the therian mammal sex determination gene *SRY*) is located on an unanchored *A. carolinensis* scaffold; these genes are thus unlikely to be the *A. carolinensis* sex determination gene.

All ten *A. carolinensis* individuals (originating from South Carolina and Tennessee) used for FISH mapping showed large pericentromeric inversions in one or more of chromosomes 1–4, with no correlation between different chromosomal inversions or with the sex of the lizard (see Supplementary Note, Supplementary Table 11 and Supplementary Fig. 6).

A total of 17,472 protein-coding genes and 2,924 RNA genes were predicted from the *A. carolinensis* genome assembly (Ensembl release 56, September 2009). We built a phylogeny for all *A. carolinensis* genes and their homologues in eight other vertebrate species (human,
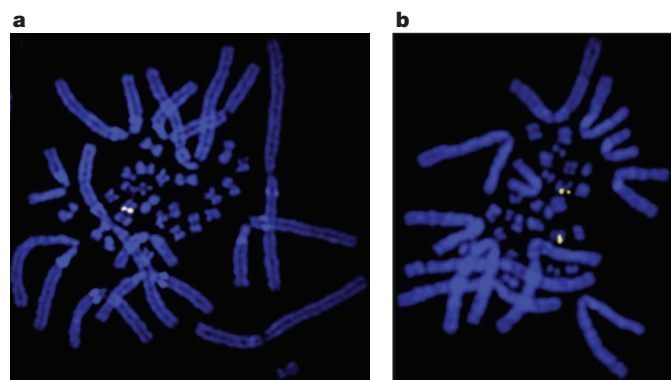
Figure 4 | The *A. carolinensis* genome contains a newly discovered X chromosome. a, b, The X chromosome, a microchromosome, is found in one copy in male *A. carolinensis* (**a**) and in two copies in females (**b**). The BAC 206M13 (CHORI-318 BAC library) is hybridized to the p arm of the X chromosome using FISH in both male and female metaphase spreads. 206M13 and ten other BACs showed this sex-specific pattern in cells derived from five male and five female individuals. Original magnification, ×1,000.

mouse, dog, opossum, platypus, chicken, zebra finch and pufferfish), allowing us to identify a conservative set of 3,994 one-to-one orthologues, that is, genes that have not been duplicated or deleted in any of these vertebrates since their last common ancestor. These gene phylogenies were also used to identify genes that arose by duplication in the lizard lineage after the split with the avian lineage and, separately, those that were lost in the mammalian lineage after the mammal–reptile split (Fig. 1, Supplementary Note, Supplementary Fig. 7 and Supplementary Table 12).

We found 11 *A. carolinensis* opsin genes that have no mammalian orthologues (but have orthologues in invertebrates, fishes and frog), and thus seem to have been lost during mammalian evolution (Supplementary Table 13). The large repertoire of opsins may contribute to the excellent colour vision of anoles—including the ability to see in the ultraviolet range—and also may contribute to their hyperdiversity by allowing the evolution of diverse, species-specific colouration of the dewlap, which has an important role in sexual selection and species recognition[11]. Similarly, olfactory receptor and β-keratin genes are highly duplicated in *A. carolinensis* (Supplementary Note and Supplementary Fig. 9).

Many reptiles, including green anoles, differ from placental mammals in being oviparous (laying eggs). Vivipary in placental mammals is a derived state, reflected in their loss of some egg-related genes. We used mass spectrometry to identify proteins present in the immature *A. carolinensis* egg, as most egg proteins are produced in the mother's body and then transported into the immature egg. We found that in contrast with mammals, reptiles have lineage-specific gene duplications, including in vitellogenins (VTGs), apovitellenin-1, ovomucin-α and three homologues of ovocalyxin-36, a chicken eggshell matrix protein.

Our results show rapid evolution of egg protein genes among amniotes. Specifically, we found proteins from 276 *A. carolinensis* genes in immature *A. carolinensis* eggs (Supplementary Tables 14 and 15), of which only 50 have been confirmed to be present in chicken eggs by mass spectrometry[22,23]. These genes include VTGs, a lysozyme, vitelline membrane outer layer protein 1 (*VMO1*) paralogues, protease inhibitors, natterin and nothepsin. By aligning genes that are one-to-one orthologues in *A. carolinensis* and chicken, we found that egg proteins evolve significantly more rapidly than non-egg proteins (mean dN/dS values (ratio of the rate of non-synonymous substitutions to the rate of synonymous substitutions) of 0.186 and 0.135, respectively; $P = 1.2 \times 10^{-5}$), which reflects reduced purifying selection and/or more frequent episodes of adaptive evolution.

Using multiple vertebrate genome sequences, we identified three *VMO1* paralogues (which we name α, β and γ) that we infer to have been present in the last common ancestor of all reptiles and mammals. Whereas at least one of *VMO1*-α, *VMO1*-β and *VMO1*-γ has been lost in all other amniote genomes, the *A. carolinensis* genome contains representatives of all three paralogues. Moreover, the *A. carolinensis*-specific *VMO1*-α family has grown to 13 members and has experienced positive selection of amino acid substitutions within a negatively charged, probably substrate-binding cavity; changes that, presumably, modify its lysozyme-like transferase activity (Supplementary Note, Supplementary Fig. 8 and Supplementary Tables 16 and 17).

The extensive and active repeat repertoire of *A. carolinensis* has allowed us to discover the origin of several mammalian conserved elements. Through the process of exaptation (a major change in function of a sequence during evolution), certain mobile elements that were active in the amniote ancestor have become conserved, and presumably functional, in mammals, while remaining active mobile elements in *A. carolinensis*. The origin of these conserved mammalian sequences in mobile elements was not recognizable without comparison to a distant and repeat-rich genome sequence[24]. We identified 96 such exapted elements (see Supplementary Table 18) in the human genome tracing back to mobile elements present in the amniote ancestor that are still present in *A. carolinensis*, particularly the CR1, L2 and gypsy families.

Although most exapted elements are non-coding and probably serve a regulatory function, we also identified a protein-coding exon that was exapted from an L2-like LINE, now constituting exon 2 in a mammal-specific N-terminal region of the *MIER1* (mesoderm induction early response 1) protein. This exon is highly conserved across 29 mammals and therefore probably represents a mammalian innovation since the amniote ancestor.

GO terms associated with the transcription start site closest to each exapted element in the human genome show enrichment for neurodevelopmental genes (see Methods), with "ephrin receptor binding", "nervous system development" and "synaptic transmission" being strongly enriched (all *P* values $< 5 \times 10^{-3}$). These enrichments are consistent with adaptive changes in neurodevelopment occurring during the emergence of mammals.

Anolis lizards are a textbook case of adaptive radiation, having diversified independently on each island in the Greater Antilles and throughout the Neotropics, producing a wide variety of ecologically and morphologically differentiated species, with as many as 15 found at a single locality[11]. Although anoles are widely used as a model system for phylogenetic comparative studies, it has been difficult to determine the evolutionary relationships among major anole clades owing to rapid evolutionary radiations associated with access to new dimensions of ecological opportunity. Successfully resolving the relatively short branching events associated with such a radiation requires a wealth of data from loci evolving at an appropriate rate.

We used the genome sequence of *A. carolinensis* to develop a new phylogenomic data set comprised of 20 kb of sequence data sampled from across the genomes of 93 species of anoles (Supplementary Tables 19 and 20). Analyses of this data set infer a well-supported phylogeny that reinforces and clarifies the adaptive and biogeographic history of anoles (Fig. 5, details in Supplementary Fig. 10). First, our phylogenomic analysis reaffirms previous molecular and morphological studies indicating that similar anole habitat specialists have
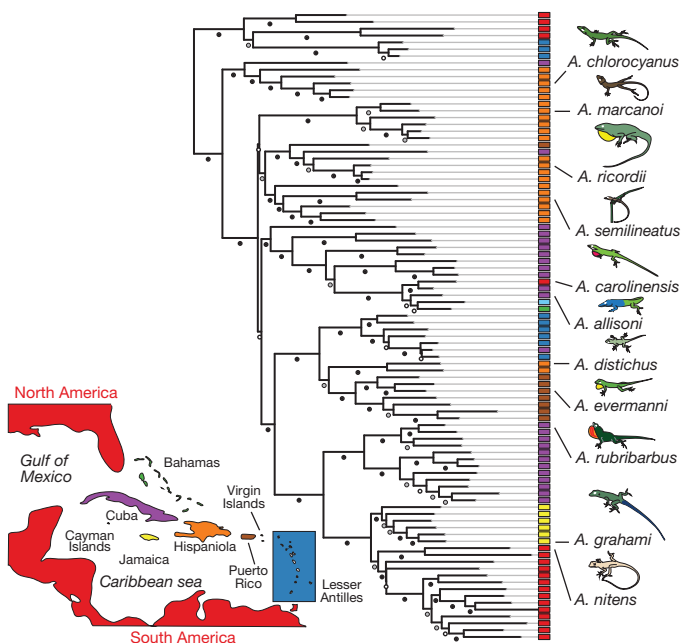
Figure 5 | A phylogeny of 93 *Anolis* species clarifies the biogeographic history of anoles. *Anolis* ecomorphs derive from convergent evolution and not from frequent inter-island migration. Using conserved primer pairs distributed across the genome of *A. carolinensis*, we obtain sequences from 46 genomically diverse loci evolving at a range of evolutionary rates and representing both protein-coding and non-coding regions. Maximum likelihood analyses of this new data set of 20 kb aligned nucleotides infer nearly all previously established anole relationships while also partially resolving the basal relationships that have plagued previous studies. Open circles indicate bootstrap (bs) values <70; grey-shaded circles, 70< bs <95; filled circles, bs >95.

evolved independently on each of the four large Greater Antillean islands. Second, our analyses suggest a complex biogeographic scenario involving a limited number of dispersal events between islands and extensive *in situ* diversification within islands. The closest relatives of *Anolis* occur on the mainland and the phylogeny confirms the existence of two colonizations, one into the southern Lesser Antilles and the second producing the diverse adaptive radiations throughout the rest of the Caribbean. Within this latter clade, anoles initially diversified primarily on the two larger Greater Antillean islands (although Puerto Rico also seems to have been involved) before subsequently undergoing secondary radiations on all of the islands and eventually returning to the mainland, where this back-colonization has produced an extensive evolutionary radiation. The phylogeny also indicates that very few inter-island dispersal events occurred in Greater Antillean evolution. Rather, the Greater Antillean faunas, renowned for the extent to which the same ecomorphs are found on each island, are primarily the result of convergent evolution[25].

The genome sequence of *A. carolinensis* allows a deeper understanding of amniote evolution. Filling this important reptilian node with a sequenced genome has revealed derived states in each major amniote branch and has helped to illuminate the amniote ancestor. However, the tree of sequenced reptilian genomes is still extremely sparse, and the sequencing of additional non-avian reptiles would be necessary to fully understand how typical *A. carolinensis* and the sequenced bird genomes are of the entire reptile clade.

In addition to the utility of the *A. carolinensis* genome sequence as a representative of non-avian reptiles, *Anolis* species are a unique resource for the study of adaptive radiation and convergent evolution. With their invasions of and subsequent radiations on Caribbean islands, anoles provide a terrestrial analogue to stickleback and cichlid fish, which underwent adaptive evolution in separate aquatic environments. Just as genomic research in sticklebacks has deepened the study of aquatic ecological speciation, a large-scale genomic phylogenetic survey of the Caribbean anoles would be an opportunity for detailed study of adaptive evolution in a land animal[26]; in particular because anole genomes contain large numbers of active mobile elements that we speculate could form substrates for exaptation of novel regulatory elements.

METHODS SUMMARY

A full description of methods, including sample collection, sequencing, assembly, anchoring, mass spectrometry and all sequence analysis, can be found in Supplementary Information. All animal experiments were approved by the MIT Committee for Animal Care.

1. Sumida, S. S. & Martin, K. L. M. *Amniote Origins: Completing the Transition to Land* (Academic Press, 1997).
2. Hillier, L. W. *et al.* Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432,** 695–716 (2004).
3. Warren, W. C. *et al.* The genome of a songbird. *Nature* **464,** 757–762 (2010).
4. Dalloul, R. A. *et al.* Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol.* **8,** e1000475 (2010).
5. Fujita, M. K., Edwards, S. V. & Ponting, C. P. Unexpected homogeneity in GC content and divergent isochore structure in the *Anolis* genome. *Genome Biol. Evol.* doi:10.1093/gbe/evr072 (in the press).
6. Benton, M. J. & Donoghue, P. C. Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24,** 26–53 (2007).
7. Axelsson, E., Webster, M. T., Smith, N. G. C., Burt, D. W. & Ellegren, H. Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. *Genome Res.* **15,** 120–125 (2005).
8. Burt, D. W. Origin and evolution of avian microchromosomes. *Cytogenet. Genome Res.* **96,** 97–112 (2002).
9. Potrzebowski, L. *et al.* Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. *PLoS Biol.* **6,** e80 (2008).
10. Organ, C. L. & Janes, D. E. Evolution of sex chromosomes in Sauropsida. *Integr. Comp. Biol.* **48,** 512–519 (2008).
11. Losos, J. B. *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles* (Univ. of California Press, 2009).
12. Gorman, G. C. in *Cytotaxonomy and Vertebrate Evolution* (eds Chiarelli, A. B. & Capanna, E.) Ch. 12 349–424 (Academic, 1973).
13. Hellsten, U. *et al.* The genome of the Western clawed frog *Xenopus* tropicalis. *Science* **328,** 633–636 (2010).
14. Bininda-Emonds, O. R. *et al.* The delayed rise of present-day mammals. *Nature* **446,** 507–512 (2007).
15. Shedlock, A. M. *et al.* Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proc. Natl Acad. Sci. USA* **104,** 2767–2772 (2007).
16. Kordis, D. Transposable elements in reptilian and avian (sauropsida) genomes. *Cytogenet. Genome Res.* **127,** 94–111 (2009).
17. Novick, P. A., Basta, H., Floumanhaft, M., McClure, M. A. & Boissinot, S. The evolutionary dynamics of autonomous non-LTR retrotransposons in the lizard *Anolis carolinensis* shows more similarity to fish than mammals. *Mol. Biol. Evol.* **26,** 1811–1822 (2009).
18. Novick, P. A., Smith, J. D., Floumanhaft, M., Ray, D. A. & Boissinot, S. The evolution and diversity of DNA transposons in the genome of the lizard *Anolis carolinensis*. *Genome Biol. Evol.* **3,** 1–14 (2010).
19. Furano, A. V., Duvernell, D. D. & Boissinot, S. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends Genet.* **20,** 9–14 (2004).
20. Duret, L., Eyre-Walker, A. & Galtier, N. A new perspective on isochore evolution. *Gene* **385,** 71–74 (2006).
21. Viets, B. E., Ewert, M. A., Talent, L. G. & Nelson, C. E. Sex-determining mechanisms in squamate reptiles. *J. Exp. Zool.* **270,** 45–56 (1994).
22. Mann, K. The chicken egg white proteome. *Proteomics* **7,** 3558–3568 (2007).
23. Farinazzo, A. *et al.* Chicken egg yolk cytoplasmic proteome, mined via combinatorial peptide ligand libraries. *J. Chromatogr. A* **1216,** 1241–1252 (2009).
24. Lowe, C. B., Bejerano, G., Salama, S. R. & Haussler, D. Endangered species hold clues to human evolution. *J. Hered.* **101,** 437–447 (2010).
25. Williams, E. E. in *Lizard Ecology: Studies of a Model Organism* (eds Huey, R. B., Pianka, E. R. & Schoener, T. W.) (Harvard Univ. Press, 1983).
26. Colosimo, P. F. *et al.* Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307,** 1928–1933 (2005).

Author Contributions J.A., F.D. and K.L.-T. planned and oversaw the project. M.G., D. Heiman and S.Y. assembled the genome. B.F.t.H., M.Y.K. and P.J.d.J. constructed the BAC library. T.C.G. and J.W. provided tissues for sequencing libraries and FISH analysis. M.B., C.W. and D. Heiman anchored the genome. T.A.C. and D.D.P. assembled the mitochondrial genome. J.K.C. and Z.S. constructed the cDNA library. S.S. and A.Z. annotated the genome. L.K., A.H. and C.P.P. performed the gene repertoire analysis. T.S. aided egg protein experimental design. J.D.J. and S.E.P. performed egg mass spectrometry. M.G. performed genome synteny analysis. E.M. performed segmental duplication analysis. C.W. and M.B. discovered the sex chromosomes and the pericentromeric inversions. P.R. performed the microchromosome and GC analysis. M.K.F. and C.P.P. participated in microchromosome and GC data interpretation. D.A.R. constructed the repeat library. D.A.R., S.B., P.A.N., A.M.S., J.D.S. and C.B. performed the repeat analysis. M.G., J.B.L., R.E.G., S.P., K.d.Q. and R.S. participated in phylogeny data collection. R.E.G., S.P., K.d.Q. and R.S. participated in phylogeny analysis. All authors participated in data discussion and interpretation. J.A., F.D., C.B.L., R.G., D.A.R., S.V.E., C.J.S., J.B.L., E.S.L., M.B., C.P.P. and K.L.-T. wrote the paper with input from other authors.

Author Information The *A. carolinensis* whole-genome shotgun project has been deposited in GenBank under the project accession AAWZ00000000.2. All phylogeny sequence data can be found at http://purl.org/phylo/treebase/phylows/study/ TB2:S11713. Reprints and permissions information is available at www.nature.com/ reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.A. (jalfoldi@broadinstitute.org) or K.L.-T. (kersli@broadinstitute.org).

Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity

Eric M. Kofoed[1] & Russell E. Vance[1]

Inflammasomes are a family of cytosolic multiprotein complexes that initiate innate immune responses to pathogenic microbes by activating the caspase 1 protease[1,2]. Although genetic data support a critical role for inflammasomes in immune defence and inflammatory diseases[3], the molecular basis by which individual inflammasomes respond to specific stimuli remains poorly understood. The inflammasome that contains the NLRC4 (NLR family, CARD domain containing 4) protein was previously shown to be activated in response to two distinct bacterial proteins, flagellin[4,5] and PrgJ[6], a conserved component of pathogen-associated type III secretion systems. However, direct binding between NLRC4 and flagellin or PrgJ has never been demonstrated. A homologue of NLRC4, NAIP5 (NLR family, apoptosis inhibitory protein 5), has been implicated in activation of NLRC4 (refs 7–11), but is widely assumed to have only an auxiliary role[1,2], as NAIP5 is often dispensable for NLRC4 activation[7,8]. However, *Naip5* is a member of a small multigene family[12], raising the possibility of redundancy and functional specialization among *Naip* genes. Here we show in mice that different NAIP paralogues determine the specificity of the NLRC4 inflammasome for distinct bacterial ligands. In particular, we found that activation of endogenous NLRC4 by bacterial PrgJ requires NAIP2, a previously uncharacterized member of the NAIP gene family, whereas NAIP5 and NAIP6 activate NLRC4 specifically in response to bacterial flagellin. We dissected the biochemical mechanism underlying the requirement for NAIP proteins by use of a reconstituted NLRC4 inflammasome system. We found that NAIP proteins control ligand-dependent oligomerization of NLRC4 and that the NAIP2–NLRC4 complex physically associates with PrgJ but not flagellin, whereas NAIP5–NLRC4 associates with flagellin but not PrgJ. Our results identify NAIPs as immune sensor proteins and provide biochemical evidence for a simple receptor–ligand model for activation of the NAIP–NLRC4 inflammasomes.

A fundamental question in immunology is how host defence is initiated in response to specific microbial ligands. The inflammasome containing the NLRC4 protein activates caspase 1 (CASP1) in response to the carboxy terminus of bacterial flagellin[6,7], as well as in response to the inner rod protein of the type III secretion systems of diverse bacterial species (for example, PrgJ of *Salmonella enterica* serovar Typhimurium)[6]. Activated CASP1 processes interleukin (IL)-1β and IL-18 inflammatory cytokines and induces a rapid and inflammatory host cell death called pyroptosis[13]. In certain cases, NLRC4 activation requires NAIP5, as *Naip5*[−/−] mice fail to activate NLRC4 or CASP1 in response to infection with *Legionella pneumophila* or in response to the C terminus of flagellin[7,8]. Interestingly, however, NAIP5 is not essential for NLRC4 activation in response to *S. enterica* Typhimurium or PrgJ[7,8].

In addition to *Naip5*, C57BL/6 mice express three other *Naip* genes (*Naip1*, *Naip2* and *Naip6*), the functions of which remain unknown[12]. We hypothesized that each NAIP paralogue may have evolved to be specific for a unique bacterial ligand. We first focused on NAIP2, as it appeared to be highly expressed in C57BL/6 mice[14]. We used specific short hairpin RNAs (shRNAs) to knock down *Naip2* expression in primary bone-marrow-derived macrophages. ShRNA1 and shRNA2 specifically reduced NAIP2 protein levels without targeting other NAIP paralogues, whereas empty vector, shRNA3 or a scrambled control shRNA had little effect on NAIP2 protein levels (Supplementary Fig. 1a, b). Macrophages expressing these shRNAs were then infected with flagellin-deficient *Listeria* strains that inducibly express PrgJ (*Listeria*-PrgJ) or flagellin (*Listeria*-FlaA)[8]. A *Listeria*-based system was chosen because it is an efficient means for delivering PrgJ to macrophages[8], and because it allows for controlled comparisons of PrgJ and FlaA within a single experimental system. Notably, knockdown of *Naip2* prevented pyroptosis and CASP1 activation by *Listeria*-PrgJ (Fig. 1a–c). By contrast, *Naip2* knockdown did not affect inflammasome activation by *Listeria*-FlaA (Fig. 1b, c) or *L. pneumophila*, which expresses flagellin but not PrgJ (Supplementary Fig. 1c). Instead, flagellin-dependent inflammasome activation depended on *Naip5*, as previously shown[7–11]. Inflammasome activation by wild-type *Salmonella*, which encodes both flagellin and PrgJ, was not significantly affected by *Naip2* knockdown (Fig. 1d, e). However, knockdown of *Naip2* in *Naip5*[−/−] macrophages significantly reduced or abolished inflammasome activation by wild-type *Salmonella* (Fig. 1d, e), indicating that both NAIP2 and NAIP5 recognize *Salmonella*. Interestingly, inflammasome activation by flagellin-deficient (FliC[−] FljB[−]) *Salmonella*, which still express PrgJ, depended entirely on *Naip2* (Fig. 1d, e). Taken together, these data indicate that *Naip2* is specifically required for activation of the NLRC4 inflammasome by PrgJ, in contrast to *Naip5*, which seems to be specifically required for NLRC4 activation by flagellin.

Biochemical analysis of the inflammasome in macrophages is complicated by the expression of multiple NAIP proteins and by their low expression levels. We therefore decided to reconstitute the NLRC4 inflammasome in non-immune 293T cells, which do not express NLRC4 or NAIPs, so that the functions of individual NAIP proteins could be analysed. 293T cells transiently transfected with green fluorescent protein (GFP)-marked vectors encoding wild-type NLRC4, NAIP5 and CASP1 did not exhibit significant spontaneous inflammasome activation, and instead, most cells expressed GFP (Fig. 2a). However, when flagellin (FlaA) from *L. pneumophila* was co-expressed with NLRC4, NAIP5 and CASP1, we observed a significant loss of GFP[high] cells and an increase in the number of dead (7AAD[+]) cells (Fig. 2a). This result was highly reminiscent of flagellin-dependent activation of the endogenous NAIP5–NLRC4 inflammasome in macrophages, which also results in a rapid CASP1-dependent cell death, loss of membrane integrity, and release of cytosolic contents and GFP[7]. Similar to the genetic requirement for *Nlrc4*, *Naip5* and *Casp1* in macrophages[4,5,7–11,15], we found that NAIP5, NLRC4, catalytically active CASP1, and FlaA are all required to trigger cell death and loss of membrane integrity/GFP in reconstituted 293T cells (Fig. 2b, c). The reconstituted NAIP5–NLRC4 inflammasome also recapitulated the ability of native inflammasomes to process CASP1 and IL-1β in response to cytosolic flagellin (Supplementary Fig. 2). Consistent with a lack of a role for NAIP5 in recognition of PrgJ by macrophages[8], the reconstituted NAIP5–NLRC4 inflammasome did not respond to PrgJ (Fig. 2d, e). By

[1]Division of Immunology and Pathogenesis, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA.
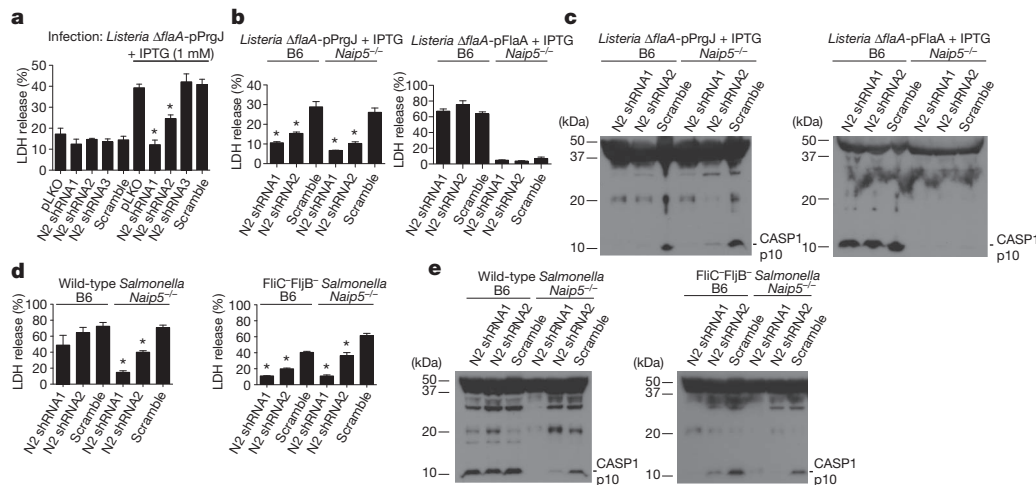
Figure 1 | NAIP2 is required in macrophages for inflammasome activation in response to PrgJ. a–c, Primary bone-marrow-derived macrophages expressing shRNAs targeting *Naip2* (N2) (or controls) were infected with flagellin-deficient *L. monocytogenes* (multiplicity of infection = 5) expressing a secreted ActA100–PrgJ (pPrgJ) or ActA100–FlaA (pFlaA) fusion protein under IPTG-inducible control. Cell death (±s.d.) was measured in triplicate by LDH release 6 h after infection (**a, b**), or active CASP1 (p10) was measured by western blotting of cell supernatants (**c**). **d, e,** NAIP2 knockdown cells were infected with wild-type or flagellin-deficient (FliC⁻FljB⁻) *S. enterica* Typhimurium and inflammasome activation was measured by LDH release (±s.d.) at 3 h after infection (**d**) or CASP1 processing (**e**). Data shown are representative of two (**c, e**) or three (**a, b, d**) independent experiments. *P < 0.02 as compared to scramble (Student's *t*-test, two-tailed).
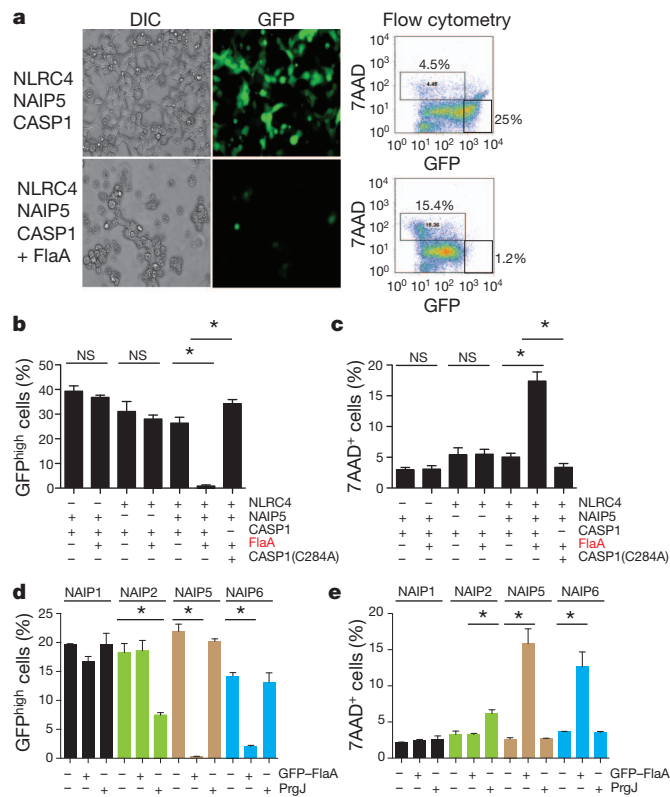
Figure 2 | Reconstitution of the NAIP5–NLRC4 inflammasome in 293T cells. a, GFP-marked expression vectors encoding NLRC4, NAIP5, CASP1 and/or flagellin (FlaA) were transiently transfected into 293T cells. Cells were imaged for differential interference contrast (DIC) and GFP fluorescence 48 h later. Dead cells were stained with 7AAD. **b, c,** GFP^high cells (**b**) and 7AAD⁺ cells (**c**) were quantified (±s.d.) as in **a**, but with specific expression vectors omitted from the transfection as indicated. CASP1(C284A) is a catalytically dead mutant. **d, e,** 293T cells were transfected as indicated and analysed as above. Data shown (±s.d.) are representative of at least three independent experiments. *P < 0.02 (Student's *t*-test, two tailed); NS, not significant.

contrast, a reconstituted NAIP2–NLRC4 inflammasome responded specifically to PrgJ but not flagellin (Fig. 2d, e). Taken together, we conclude that we have successfully reconstituted NAIP2–NLRC4 and NAIP5–NLRC4 inflammasomes that exhibit all the known requirements and specificities of the native inflammasomes.

It is believed that activated inflammasomes assemble into high-molecular-mass multiprotein complexes[16], but this has not been demonstrated for the NLRC4 inflammasome. To visualize inflammasome assembly, 293T cells were transfected with NAIP5, NLRC4 and FlaA in various combinations, but CASP1 was omitted so that cell death and loss of cellular contents (and assembled inflammasomes) would not occur. Digitonin-solubilized cell lysates were resolved on blue native (BN) polyacrylamide gel electrophoresis (PAGE) gels[17]. A marked shift of NLRC4 from a monomer (~120 kDa) to an oligomeric complex (~1,000 kDa) was seen in the presence of NAIP5 and FlaA. NAIP5 was also contained within the high-molecular-mass oligomeric complex (Fig. 3a). The association of NAIP5 and NLRC4 in the same complex was validated by co-immunoprecipitation (Supplementary Fig. 3)[11,18]. NLRC4 oligomerization was induced by either untagged FlaA or a GFP–FlaA fusion protein (Fig. 3a), both of which activate NLRC4–CASP1. Importantly, assembly of the NLRC4 inflammasome required FlaA (Supplementary Fig. 4a) and was not observed in the absence of NAIP5 (Fig. 3a), indicating that a biochemical function of NAIP5 is to promote NLRC4 oligomerization.

Despite strong genetic evidence that NLR proteins, such as NAIP5 and NLRC4, function as microbial 'sensors', there is no biochemical evidence that NLRs interact directly with microbial ligands. In fact, some studies of the NLRP3 inflammasome[19–21], as well as analyses of analogous proteins from plants[22], suggest that at least some NLRs recognize pathogens indirectly. To determine if the oligomerized NAIP5–NLRC4 complex also contains flagellin, we subjected samples separated in the first dimension by native PAGE to a second dimension of SDS–PAGE. To facilitate detection of flagellin, we used a 6×Myc-tagged flagellin, which activates the inflammasome identically to native flagellin (data not shown). This approach revealed that FlaA was indeed present in a high-molecular-mass complex, along with NAIP5 and NLRC4 (Fig. 3b). NAIP5 exhibited a weak flagellin-dependent mobility shift in the absence of NLRC4 (Supplementary Fig. 4b), indicating that NLRC4 is not essential for flagellin recognition, although formation/stabilization of the oligomerized complex seems to be significantly enhanced by NLRC4. FlaA expressed alone was present in cell extracts
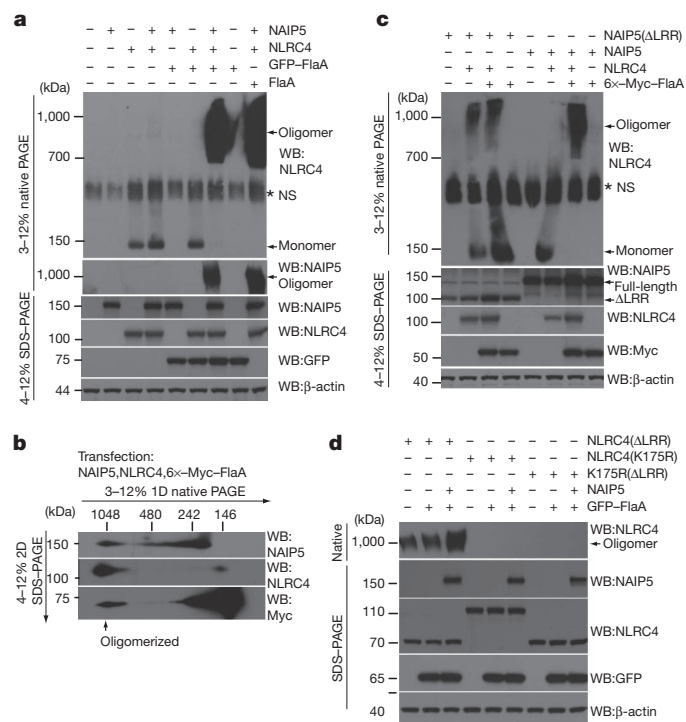
Figure 3 | NAIP5 is required for formation of a hetero-oligomeric complex that contains NLRC4, NAIP5 and flagellin. a, 293T cells were transfected as indicated, followed by analysis by blue native PAGE or SDS–PAGE, and western blotting. *NS, nonspecific band. **b**, 293T cells were transfected as indicated and lysates were separated by a first dimension of blue native PAGE followed by a second dimension of SDS–PAGE. **c, d**, 293T cells were transfected as indicated and samples were processed and analysed as in **a**. Data shown are representative of at least three independent experiments.

only as a monomer (Supplementary Fig. 4c). Taken together, these observations provide evidence for a simple receptor–ligand model of NAIP5–NLRC4 activation by flagellin.

Consistent with the autoinhibitory function of the leucine-rich repeats (LRRs) in other NLRs, we found that NAIP5(ΔLRR) and NLRC4(ΔLRR) constitutively activated CASP1-dependent cell death, independent of the presence of flagellin (Supplementary Fig. 5). Interestingly, NLRC4(ΔLRR) was able to activate CASP1 in the absence of NAIP5, whereas constitutively active NAIP5(ΔLRR) required wild-type NLRC4 to activate CASP1. This result suggests that NAIP5 functions upstream of NLRC4. Indeed, NAIP5(ΔLRR) was able to induce the oligomerization of wild-type NLRC4 (Fig. 3c), whereas the spontaneous oligomerization of NLRC4(ΔLRR) did not require NAIP5 (Fig. 3d). Spontaneous oligomerization of NLRC4(ΔLRR) did require the nucleotide-binding domain (NBD) of NLRC4, as a K175R mutation previously shown to disrupt NBD function[23] abolished NLRC4(ΔLRR) auto-oligomerization (Fig. 3d). The ability of NAIP5 to induce oligomerization of NLRC4 in response to flagellin required both the NBD and amino-terminal BIRs of NAIP5, but did not require the N-terminal CARD of NLRC4 (Supplementary Fig. 4d, e), whereas functional CASP1 activation required all of these domains (Supplementary Figs 5 and 6). Taken together, these data are indicative of a working model (Supplementary Fig. 7) in which NAIP5 is activated by flagellin and induces downstream NLRC4 oligomerization and CASP1 activation.

Consistent with a specific role for NAIP2 in recognition of PrgJ, we found that PrgJ did not induce the oligomerization of NAIP5–NLRC4 (Fig. 4a) but did induce oligomerization of NAIP2–NLRC4. Oligomerization of NLRC4 did not occur when co-expressed with NAIP2 alone or with NAIP2 and FlaA (Fig. 4b). Interestingly, NAIP6 resembled NAIP5 and supported NLRC4 oligomerization in response to FlaA but
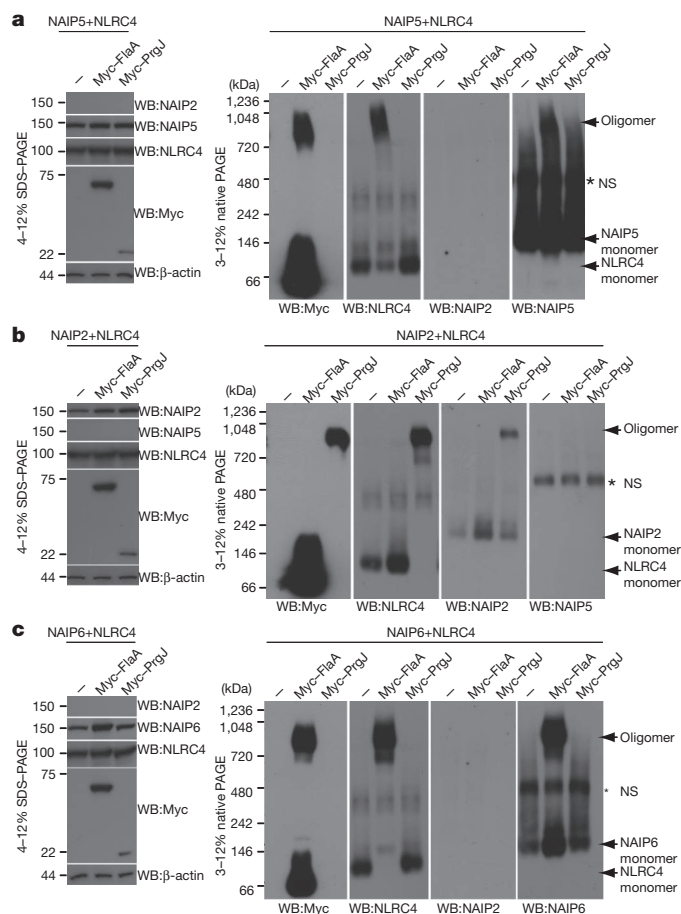
Figure 4 | NAIP paralogues confer specificity to the NLRC4 inflammasome. a, 293T cells were co-transfected with wild-type NAIP5 and NLRC4, alone or in combination with 6×-Myc–FlaA or 6×-Myc–PrgJ followed by blue native PAGE 48 h later. *NS, nonspecific band. Whole-cell lysates were also separated by conventional 4–12% SDS–PAGE to control for expression of each transfected gene construct (left panel). **b**, 293T cells were transfected with wild-type NAIP2 and NLRC4 and analysed as in **a**. **c**, 293T cells were transfected with wild-type NAIP6 and NLRC4, and analysed as in **a**. Data shown are representative of at least three independent experiments.

not PrgJ (Fig. 4c), perhaps providing an explanation for the previously puzzling observation that $Naip5^{-/-}$ cells can respond to high levels of flagellin[7]. In contrast, NAIP1 is an 'orphan' NAIP because it responded neither to PrgJ nor flagellin (Fig. 2d, e and Supplementary Fig. 8).

Our results demonstrate that the ability of the NLRC4 inflammasome to assemble and functionally activate CASP1 in response to specific bacterial ligands is dictated by NAIP family members. The most parsimonious model to account for our results is that NAIP proteins function as direct receptors for bacterial ligands (Supplementary Fig. 7). Although NLRC4 was previously suspected to be the cytosolic flagellin sensor[1,2], we hypothesize that a main function of NLRC4 may instead be to serve as an adaptor, downstream of NAIP proteins, to recruit CASP1 via a CARD–CARD interaction. NLRC4 may also have an important function in ligand binding or in stabilizing NAIP–NLRC4–ligand complexes, but the specificity of the complexes for particular ligands seems to be controlled by NAIP proteins.

The number and sequence of *Naip* paralogues vary significantly among inbred mouse strains, and have been suggested to be evolving rapidly[24]. Indeed, the murine *Naip* locus was originally identified by a forward genetic approach which took advantage of the widely varying susceptibility of inbred mouse strains to *L. pneumophila* infection[14,25]. The single known human *NAIP* orthologue may also exist within a rapidly evolving locus[24]; our results indicate that it will be of great interest to establish the specificity of the human NAIP protein. We

propose that *Naip* gene evolution represents a fascinating example of the molecular arms race between bacteria and their hosts.

METHODS SUMMARY

***Naip2* knockdown.** Primary C57BL/6 bone marrow cells were transduced with pLKO.1-based lentivirus encoding shRNAs that specifically target *Naip2* or controls. Bone marrow cells were differentiated into macrophages by culture in media containing macrophage colony-stimulating factor (MCSF). On day 4 of culture, transduced macrophages were selected by addition of puromycin ($5\,\mu g\,ml^{-1}$). On day 8 of culture, macrophages were re-plated and infected the next day with *Listeria monocytogenes* or *S. enterica* Typhimurium expressing flagellin or PrgJ[8], and inflammasome activation was measured by assaying release of the cytosolic enzyme lactate dehydrogenase (LDH)[7] or by western blotting for processed (p10) CASP1.

Reconstituted inflammasome. The inflammasome was reconstituted by transfection of 293T cells with MSCV2.2-IRES-GFP-based expression vectors encoding various mouse (C57BL/6-derived) *Naip* genes, *Nlrc4* and caspase 1. Inflammasome oligomerization was assessed in digitonin (1%) lysates using a Bis-Tris NativePAGE system (Invitrogen) followed by western blotting.

Statistical analysis. Statistical differences were calculated with an unpaired two-tailed Student's *t*-test using GraphPad Prism 5.0b.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Latz, E. The inflammasomes: mechanisms of activation and function. *Curr. Opin. Immunol.* **22,** 28–33 (2010).
2. Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140,** 821–832 (2010).
3. Ting, J. P., Kastner, D. L. & Hoffman, H. M. CATERPILLERs, pyrin and hereditary immunological disorders. *Nature Rev. Immunol.* **6,** 183–195 (2006).
4. Franchi, L. *et al.* Cytosolic flagellin requires Ipaf for activation of caspase-1 and interleukin 1β in salmonella-infected macrophages. *Nature Immunol.* **7,** 576–582 (2006).
5. Miao, E. A. *et al.* Cytoplasmic flagellin activates caspase-1 and secretion of interleukin 1β via Ipaf. *Nature Immunol.* **7,** 569–575 (2006).
6. Miao, E. A. *et al.* Innate immune detection of the type III secretion apparatus through the NLRC4 inflammasome. *Proc. Natl Acad. Sci. USA* **107,** 3076–3080 (2010).
7. Lightfield, K. L. *et al.* Critical function for Naip5 in inflammasome activation by a conserved carboxy-terminal domain of flagellin. *Nature Immunol.* **9,** 1171–1178 (2008).
8. Lightfield, K. L. *et al.* Differential requirements for NAIP5 in activation of the NLRC4 (IPAF) inflammasome. *Infect. Immun.* **79,** 1606–1614 (2011).
9. Molofsky, A. B. *et al.* Cytosolic recognition of flagellin by mouse macrophages restricts *Legionella pneumophila* infection. *J. Exp. Med.* **203,** 1093–1104 (2006).
10. Ren, T., Zamboni, D. S., Roy, C. R., Dietrich, W. F. & Vance, R. E. Flagellin-deficient *Legionella* mutants evade caspase-1- and Naip5-mediated macrophage immunity. *PLoS Pathog.* **2,** e18 (2006).
11. Zamboni, D. S. *et al.* The Birc1e cytosolic pattern-recognition receptor contributes to the detection and control of *Legionella pneumophila* infection. *Nature Immunol.* **7,** 318–325 (2006).
12. Growney, J. D. & Dietrich, W. F. High-resolution genetic and physical map of the Lgn1 interval in C57BL/6J implicates Naip2 or Naip5 in *Legionella pneumophila* pathogenesis. *Genome Res.* **10,** 1158–1171 (2000).
13. Bergsbaken, T., Fink, S. L. & Cookson, B. T. Pyroptosis: host cell death and inflammation. *Nature Rev. Microbiol.* **7,** 99–109 (2009).
14. Wright, E. K. *et al.* Naip5 affects host susceptibility to the intracellular pathogen *Legionella pneumophila. Curr. Biol.* **13,** 27–36 (2003).
15. Broz, P., von Moltke, J., Jones, J. W., Vance, R. E. & Monack, D. M. Differential requirement for Caspase-1 autoproteolysis in pathogen-induced cell death and cytokine processing. *Cell Host Microbe* **8,** 471–483 (2010).
16. Martinon, F., Burns, K. & Tschopp, J. The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-β. *Mol. Cell* **10,** 417–426 (2002).
17. Schagger, H., Cramer, W. A. & von Jagow, G. Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Anal. Biochem.* **217,** 220–230 (1994).
18. Damiano, J. S., Oliveira, V., Welsh, K. & Reed, J. C. Heterotypic interactions among NACHT domains: implications for regulation of innate immune responses. *Biochem. J.* **381,** 213–219 (2004).
19. Hornung, V. *et al.* Silica crystals and aluminum salts activate the NALP3 inflammasome through phagosomal destabilization. *Nature Immunol.* **9,** 847–856 (2008).
20. Zhou, R., Tardivel, A., Thorens, B., Choi, I. & Tschopp, J. Thioredoxin-interacting protein links oxidative stress to inflammasome activation. *Nature Immunol.* **11,** 136–140 (2010).
21. Zhou, R., Yazdi, A. S., Menu, P. & Tschopp, J. A role for mitochondria in NLRP3 inflammasome activation. *Nature* **469,** 221–225 (2011).
22. Chisholm, S. T., Coaker, G., Day, B. & Staskawicz, B. J. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* **124,** 803–814 (2006).
23. Lu, C. *et al.* Nucleotide binding to CARD12 and its role in CARD12-mediated caspase-1 activation. *Biochem. Biophys. Res. Commun.* **331,** 1114–1119 (2005).
24. Romanish, M. T., Lock, W. M., de Lagemaat, L. N., Dunn, C. A. & Mager, D. L. Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus *NAIP* during mammalian evolution. *PLoS Genet.* **3,** e10 (2007).
25. Diez, E. *et al.* Birc1e is the gene within the *Lgn1* locus associated with resistance to *Legionella pneumophila. Nature Genet.* **33,** 55–60 (2003).

METHODS

Mice. C57BL/6J (B6) mice were purchased from Jackson Labs and bred at UC Berkeley. *Naip5*[−/−] mice on a pure B6 background were described previously[7]. Animal experiments were approved by the UC Berkeley Animal Care and Use Committee.

Cell culture. HEK293T cells were grown in complete media (DMEM, 10% FBS, 100 units ml[−1] penicillin, 100 μg ml[−1] streptomycin, 2 mM L-glutamine). Bone marrow macrophages were differentiated from bone-marrow-derived precursor cells using macrophage colony stimulating factor as described previously[7].

Transient transfections. HEK293T cells were plated at 8×10^5 cells per well in 6-well plates, transfected the next day using Lipofectamine 2000, and collected for flow cytometric analysis 48 h later.

Measurement of cell death. Cell death of HEK293T cells was measured by flow cytometry measuring GFP and 7AAD fluorescence. Cells were stained with 7AAD (BD Pharmingen) according to the manufacturer's instructions. Death of immortalized macrophages in response to *L. pneumophila* and LFn-FlaA was measured by following release of the intracellular enzyme lactate dehydrogenase (LDH) as described previously[7]. Infection of immortalized macrophages with *L. pneumophila* was performed using a multiplicity of infection (MOI) of 2, and cell death was measured 4 h later as described previously[7]. LFn-FlaA is a recombinant 6×His-tagged fusion protein encoding the first (non-enzymatically active) 263 amino acids of lethal factor from *Bacillus anthracis* fused to full-length flagellin (FlaA) from *L. pneumophila* (J. von Moltke and R.E.V., unpublished). LFn-FlaA and *B. anthracis* protective antigen (PA) were purified from *E. coli* as described previously[26]. Endotoxin was removed from these proteins using Detoxi-Gel (Pierce) according to the manufacturer's protocol.

Western blotting and native PAGE. Blue native gel electrophoresis was performed using the Bis-Tris NativePAGE system by Invitrogen according to the manufacturer's instructions. Briefly, HEK293T cells were transfected for 48 h, washed twice with cold PBS, trypsinized for 3 min at 37 °C, re-suspended in complete media, and pelleted by centrifugation at 400*g* for 5 min at 4 °C. Cell pellets were washed twice with cold PBS, followed by re-suspension in 1% digitonin native lysis buffer (50 mM BisTris, 50 mM NaCl, 10% w/v glycerol, 0.001% Ponceau S, 1% digitonin, 2 mM Na$_3$VO$_4$, 1 mM PMSF, 25 mM NaF, 1× Roche protease inhibitor cocktail (no EDTA), pH 7.2). Cell lysates were triturated and incubated on ice for 30 min, and then insoluble cell debris was pelleted by centrifugation at 16,000*g* for 30 min at 4 °C. Lysates were quantified for total protein using the BCA protein assay (Pierce), equalized for total protein, and separated by NativePAGE using the Novex BisTris gel system according to the manufacturer's instructions (Invitrogen). Native gels were soaked in 10% SDS for 5 min before transfer to PVDF membrane (Millipore) and conventional western blotting. Antibodies used were anti-mIL-1β (R&D systems), anti-caspase 1 p10 (M20) (Santa Cruz), anti-NLRC4 (gift of S. Mariathasan and V. Dixit, Genentech), anti-NAIP5(961–978)[14], anti-Myc (9E10) (Clontech), anti-β-actin (C4) (Santa Cruz), anti-GFP (JL-8) (Clontech), anti-NAIP2(33–46)[14], anti-rabbit IgG-HRP and anti-mouse IgG-HRP (GE Healthcare), anti-goat IgG-HRP (Santa Cruz). In some cases, native gels were subjected to a second dimension of electrophoresis using SDS–PAGE. A 5.7-cm slice (lane) of natively resolved gel was placed in a dish containing 1× Laemmli sample buffer (50 mM Tris-Cl (pH 6.8), 100 mM DTT, 2% (w/v) SDS, 0.1% bromophenol blue, 10% (v/v) glycerol) for 10 min, microwaved on high for 20 s, and rocked for another 5 min before loading slice into the well of a precast 2D 4–12% SDS–PAGE gel (Invitrogen). For immunoprecipitations, digitonin cell extracts (100 μg total protein) were pre-cleared with 25 μl of washed Protein-G-sepharose (GE Healthcare), and then cleared extracts were incubated with 1 μg primary antibody (or isotype controls) overnight at 4 °C, and complexes were captured the following day with 25 μl of washed Protein-G-sepharose. Bound proteins were eluted by re-suspension in Laemmli sample buffer, boiled for 5 min and separated by SDS–PAGE.

Expression constructs. NAIP5 wild-type and mutant constructs were cloned into the MSCV2.2 retroviral vector containing an IRES-GFP downstream of the multiple cloning site. Expression is driven from the viral LTR and is considerably lower than that driven by the CMV promoter (data not shown). In general, wild-type and mutant ORFs were amplified between flanking BamHI and NotI sites, and a Kozak sequence (GCCACC) was engineered to precede the start codon. The BamHI/NotI digested PCR product was cloned into complementary BglII/NotI digested MSCV2.2 vector. Wild-type NAIP5(1–1402) and NAIP5(ΔPloop) (Δ464–487) were amplified from pcDNA3 using forward (AAAAGGATCCGCCACCATGGCTGAGCATGGGGAGTCCTCCG) and reverse (AAAAGCGGCCGCTTACTCCAGGATAACAGGAGAGAATGGGAC) primers. NAIP5(Δ347) (347–1402) was generated using the forward primer (AAAAGGATCCGCCACCATGACCTTGAAGTCCTCTGCAGAAG) in combination with wild-type NAIP5 reverse primer. NAIP5(ΔLRR) (1–1039) was PCR cloned into BglII/PmeI sites of MSCV2.2 vector using the wild-type NAIP5 forward primer (see above), and

reverse primer (GTTTAAACTCAGCCACTGCTGTTGAATAAACG). Wild-type NLRC4 (1–1024) was PCR amplified from pCDNA3 using forward (AAAAGGATCCGCCACCATGAACTTTATAAGGAACAACAGACG) and reverse (TTTTGCGGCCGCTTAAGCAGTCACTAGTTTAAAGGTGCC) primers, and ligated into BglII/NotI sites of MSCV2.2-IRES-GFP. NLRC4(K175R) was generated by site-directed mutagenesis using the QuickChange protocol (Stratagene) using forward (GAGTCTGGCAAAGGGCGATCGACCCTGCTGCAG) and reverse (CTGCAGCAGGGTCGATCGCCCTTTGCCAGACTC) primers. NLRC4(ΔCARD) (89–1024) was PCR amplified from pCDNA3 using forward primer (AAAAGGATCCGCCACCATGTCTTATCAGGTCACAGAAGAAGACC) paired with the wild-type reverse primer. NLRC4(ΔLRR) (1–656) was PCR amplified from pCDNA3 using the reverse primer (TTTTGCGGCCGCTTACTTCCAGTTGAAGAACAAAGCACACAGC) in combination with the wild-type forward primer. Wild-type NAIP2 was PCR cloned from pBluescript(SK-) into the BglII/NotI sites of MSCV2.2-IRES-GFP using forward (AAAAGGATCCGCCACCATGGCAGCCCAGGGAGAAGCCGTTGAGG) and reverse (TTTTGCGGCCGCTCACTTCTGAATGACAGGAGGAATGGCACTACCC) primers. Wild-type NAIP6 was PCR cloned from pBluescript(SK-) into the BglII/NotI sites of Mscv2.2-IRES-GFP using the wild-type NAIP5 forward primer (there is 100% conservation of the first 26 nucleotides among NAIP1, NAIP5 and NAIP6) and the reverse primer (TTTTGCGGCCGCTTACTCCAGGACAACAGGAGAGAACGGGAC). Wild-type NAIP1 was PCR cloned from pBluescript(SK-) into the BglII/NotI sites of MSCV2.2-IRES-GFP using the wild-type NAIP5 forward primer, and the wild-type NAIP6 reverse primer (NAIP1 and NAIP6 share identical C-terminal nucleotides).

MSCV2.2-caspase 1 and MSCV2.2-caspase 1(C284A) have been described previously[15].

We modified the MSCV2.2 vector to contain a 6×-Myc-tag in the MCS to facilitate generation of NH$_2$-terminal fusion proteins (6×-Myc-MSCV2.2-IRES-GFP). The 6×-Myc-tag from pCS-6×-Myc-SEC24 was PCR amplified using forward (AAAAAGATCTATCGATTTAAAGCTATG) and reverse (TTTTGCGGCCGCTGG-CCGGCCTGAATTCA) primers for insertion into the BglII site of MSCV2.2. 6×-Myc-FlaA was generated by PCR amplifying full-length FlaA (*L. pneumophila*) from MSCV2.2-FlaA[7] using forward (AAAAGCGGCCGCAGCTCAAGTAATCAACACTAATGTGGC) and reverse (TTTTGTCGACTATC-GACCTAACAAAGATAATAC) primers and inserted into NotI/SalI sites. 6×-Myc-PrgJ was generated by amplifying PrgJ from MSCV2.2-PrgJ using forward (AAAAGCGGCCGCATCGATTGCAACTATTGTCCC) and reverse (TTTTGTCGACTCATGAGCGTAATAGCGTTTC) primers and insertion into NotI/SalI sites. All constructs were fully sequenced to confirm their identity. Sequencing primers used for NAIP5: MSCV2.2-F, AAGCCCTTTGTACACCCTAAGCC, MSCV2.2-R, CCTCACATTGCCAAAAGAC; NAIP5seq#1, CAGCAAAAGCACTGAACGCC; NAIP5seq#2, ATGAACAAATCCCTCGTAGC; NAIP5seq#3, TCACTCCTACCCAAGTCCAC; NAIP5seq#4, CTCAGACACACTTCACTAATGC; NAIP5seq#5, TCCCTTAGTTCCATCACACC; NAIP5seq#6, GACCCCTCTCTTTGTAGCAG; NAIP5seq#7, GAGTTTCTTGCTGCCGTGAG; NAIP5seq#8, TTAGAGGGTTGT-GGCTGGTGTC; NAIP5seq#9, CTTCACAGAGTATTGAGTTCCG; NAIP5seq#10, TTGAGTTTTCTGGACGATGC; NAIP5seq#11, GGACAACTTGCCAAACCTAC. Sequencing primers for NAIP2: NAIP2seqF1, TGGTGATGAAAAGAGTCAC; NAIP2seqF2, CTTCACAGAGTATTGAGTTCCG; NAIP2seqR1, AGCAAATGGTCAGTGCCGAG; NAIP2seqR2, ACATACTGCTGCCACGAAG; NAIP2seqR3, AATCCAGTGTTCTCCCCTCG; MSCV2.2-F and -R primers (see above). Sequencing primers for NAIP6: NAIP6seqF1, CAGAAAGCCTGTTACTGTTGAG; NAIP6seqR1, GATGGAACTAAGGGAGAGGTAG; NAIP6seqR2, TCTTGGTCTTCCTGCCTATC; MSCV2.2-F, MSCV2.2-R, NAIP5seq#7 (see above).

Generation of *Naip2* shRNA constructs. We used the lentiviral pLKO.1-TRC cloning vector (Addgene) to generate three vectors expressing three separate shRNAs targeting *Naip2* (*Naip2* shRNAs 1–3). *Naip2* shRNA1 oligos, CCGGGCCATTGCCTTTCAACCTATACTCGAGTATAGGTTGAAAGGCAATGGCTTTTG and AATTCAAAAAGCCATTGCCTTTCAACCTATACTCGAGTATAGGTTGAAAGGCAATGGC. *Naip2* shRNA2 oligos, CCGGCCATCCAGAAACCTTGTTGTTCTCGAGAACAACAAGGTTTCTGGATGGTTTTTG and AATTCAAAAACCATCCAGAAACCTTGTTGTTCTCGAGAACAACAAGGTTTCTGGATGG. *Naip2* shRNA3 oligos, CCGGCTTTCAGTCTTGAAGAGACAACTCGAGTTGTCTCTTCAAGACTGAAAGTTTTTG and AATTCAAAAACTTTCAGTCTTGAAGAGACAACTCGAGTTGTCTCTTCAAGACTGAAAG. We included pLKO.1-TRC control vector (Addgene ID#10879) or pLKO.1 scramble (Addgene ID#1864) as negative controls.

Knockdown of NAIP2 in primary macrophages. Bone marrow was collected from C57BL/6 mice on day 0, and plated into one 15-cm plate in macrophage differentiation media (see above). Lentivirus encoding *Naip2* shRNAs were generated according to the Addgene protocol. On day 2, primary bone marrow macrophages were collected, red blood cells were lysed, and cells were plated at 1×10^6 cells per

well in 6-well plates. Macrophages were spinfected with lentiviral particles at 32 °C, 90 min, 1,258g, and placed in a 32 °C incubator for 48 h. On day 4, cells were collected in cold PBS, re-plated on 10-cm plates containing fresh media containing puromycin (5 µg ml^{-1}) for selection, and placed at 37 °C. Puromycin containing media was replaced on day 6. On day 8, macrophages were collected, counted and seeded at 1×10^5 cells per well in a 96-well plate, and infected the following day. In some experiments, NAIP2 was knocked down in v-myc/v-raf immortalized bone-marrow-derived macrophages that were previously generated[15] by use of the J2 virus.

Listeria infections. Flagellin-deficient strains of *Listeria monocytogenes* 10403S were generated that express the secreted fusion protein ActAN100–PrgJ (*Listeria* Δ*flaA*-pPrgJ) or ActAN100–FlaA (*Listeria* Δ*flaA*-pFlaA), under the control of an IPTG-inducible promoter, as previously described[8]. Macrophages were spinfected in antibiotic free media at 400g for 10 min at an MOI = 5, with or without IPTG

(1 mM), and placed at 37 °C for 30 min. The media was then replaced with complete media containing gentamycin (10 µg ml^{-1}) and IPTG, and supernatants were assayed for LDH assay 5.5 h later.

Salmonella infections. *Salmonella enterica* serovar Typhimurium strain LT2 or isogenic flagellin mutant (FliC$^-$ FljB$^-$) was grown in 10 ml LB standing cultures at 37 °C overnight. The next morning, the cultures were diluted 1:100 in LB and grown for 4 h (standing culture, 37 °C). Bacteria were added to macrophages at an MOI of 10–30, followed by centrifugation at 400g for 10 min. Gentamycin (25 µg ml^{-1}) was added after 1 h to kill remaining extracellular bacteria. Caspase 1 processing or LDH release was monitored as previously described[7].

26. Krantz, B. A. *et al.* A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore. *Science* **309**, 777–781 (2005).

ARTICLE

lincRNAs act in the circuitry controlling pluripotency and differentiation

Mitchell Guttman[1,2], Julie Donaghey[1], Bryce W. Carey[2,3], Manuel Garber[1], Jennifer K. Grenier[1], Glen Munson[1], Geneva Young[1], Anne Bergstrom Lucas[4], Robert Ach[4], Laurakay Bruhn[4], Xiaoping Yang[1], Ido Amit[1], Alexander Meissner[1,5]*, Aviv Regev[1,2]*, John L. Rinn[1,5]*, David E. Root[1]* & Eric S. Lander[1,2,6]

Although thousands of large intergenic non-coding RNAs (lincRNAs) have been identified in mammals, few have been functionally characterized, leading to debate about their biological role. To address this, we performed loss-of-function studies on most lincRNAs expressed in mouse embryonic stem (ES) cells and characterized the effects on gene expression. Here we show that knockdown of lincRNAs has major consequences on gene expression patterns, comparable to knockdown of well-known ES cell regulators. Notably, lincRNAs primarily affect gene expression in *trans*. Knockdown of dozens of lincRNAs causes either exit from the pluripotent state or upregulation of lineage commitment programs. We integrate lincRNAs into the molecular circuitry of ES cells and show that lincRNA genes are regulated by key transcription factors and that lincRNA transcripts bind to multiple chromatin regulatory proteins to affect shared gene expression programs. Together, the results demonstrate that lincRNAs have key roles in the circuitry controlling ES cell state.

The mammalian genome encodes many thousands of large non-coding transcripts[1] including a class of ~3,500 lincRNAs identified using a chromatin signature of actively transcribed genes[2–4]. These lincRNA genes have been shown to have interesting properties, including clear evolutionary conservation[2–5], expression patterns correlated with various cellular processes[2,6] and binding of key transcription factors to their promoters[2,6], and the lincRNAs themselves physically associate with chromatin regulatory proteins[4,7]. Yet, it remains unclear whether the RNA transcripts themselves have biological functions[8–10]. Few have been demonstrated to have phenotypic consequences by loss-of-function experiments[6]. As a result, the functional role of lincRNA genes has been widely debated. Various proposals include that lincRNA genes act as enhancer regions, with the RNA transcript simply being an incidental by-product[8,9], that lincRNA transcripts act in *cis* to activate transcription[11], and that lincRNA transcripts can act in *trans* to repress transcription[12,13].

We therefore sought to undertake systematic loss-of-function experiments on all lincRNAs known to be expressed in mouse embryonic stem (ES) cells[2,3]. ES cells are pluripotent cells that can self-renew in culture and can give rise to cells of any of the three primary germ layers including the germ line[14]. The signalling[14], transcriptional[15–17] and chromatin[15,18–21] regulatory networks controlling pluripotency have been well characterized, providing an ideal system to determine how lincRNAs may integrate into these processes.

Here we show that knockdown of the vast majority of ES-cell-expressed lincRNAs has a strong effect on gene expression patterns in ES cells, of comparable magnitude to that seen for the well-known ES cell regulatory proteins. We identify dozens of lincRNAs that upon loss-of-function cause an exit from the pluripotent state and dozens of additional lincRNAs that, although not essential for the maintenance of pluripotency, act to repress lineage-specific gene expression programs in ES cells. We integrate the lincRNAs into the molecular circuitry of ES cells by demonstrating that most lincRNAs are directly regulated by critical pluripotency-associated transcription factors and ~30% of lincRNAs physically interact with specific chromatin regulatory proteins to affect gene expression. Together, these results demonstrate a regulatory network in ES cells whereby transcription factors directly regulate the expression of lincRNA genes, many of which can physically interact with chromatin proteins, affect gene expression programs and maintain the ES cell state.

lincRNAs affect global gene expression

To perform loss-of-function experiments, we generated five lentiviral-based short hairpin RNAs (shRNAs)[22] targeting each of the 226 lincRNAs previously identified in ES cells[2,3] (see Methods and Supplementary Table 1). These shRNAs successfully targeted 147 lincRNAs and reduced their expression by an average of ~75% compared to endogenous levels in ES cells (see Methods, Fig. 1a, Supplementary Fig. 1 and Supplementary Table 2). As positive controls, we generated shRNAs targeting ~50 genes encoding regulatory proteins, including both transcription and chromatin factors that have been shown to play critical roles in ES cell regulation[17,20,23]; validated hairpins were obtained against 40 of these genes (Supplementary Table 2). As negative controls, we performed independent infections with lentiviruses containing 27 different shRNAs with no known cellular target RNA.

We infected each shRNA into ES cells, isolated RNA after 4 days, and profiled their effects on global transcription by hybridization to genome-wide microarrays (Fig. 1a, see Methods). We used a stringent procedure to control for nonspecific effects due to viral infection, generic RNA interference (RNAi) responses, or 'off-target' effects. Expression changes were deemed significant only if they exceeded the maximum levels observed in any of the negative controls, showed a twofold change in expression compared to the negative controls, and had a low false discovery rate (FDR) assessed across all genes based on permutation tests (Fig. 1b, see Methods). This approach controls for the overall rate of nonspecific effects by estimating the number and

Figure 1 | Functional affects of lincRNAs. a, A schematic of lincRNA perturbation experiments. ES cells are infected with shRNAs, knockdown level is computed, the best hairpin is selected and profiled on expression arrays, and differential gene expression is computed relative to negative control hairpins. **b**, Example of a lincRNA knockdown. Top: genomic locus containing the lincRNA. Bottom: heat-map of the 95 genes affected by knockdown of the lincRNA, expression for control hairpins (red line) and expression for lincRNA hairpins (blue line) are shown. **c**, Distribution of number of affected genes upon knockdown of 147 lincRNAs (blue) and 40 well-known ES cell regulatory proteins (red). Points corresponding to five specific ES cell regulatory proteins are marked.

magnitude of observed effects in the negative control hairpins, where all effects are nonspecific.

For 137 of the 147 lincRNAs (93%), knockdown caused a significant impact on gene expression (Supplementary Table 3), with an average of 175 protein-coding transcripts affected (range: 20–936) (Fig. 1c, Supplementary Fig. 2 and Supplementary Table 4). These results were similar to those obtained upon knockdown of the 40 well-studied ES cell regulatory proteins: 38 (95%) showed significant effects on gene expression, with an average of 207 genes affected (range: 28 (for Dnmt3l) to 1,187 (for Oct4)) (Fig. 1c, Supplementary Fig. 2 and Supplementary Table 4). Although some individual lincRNAs have been found to lead primarily to gene repression[12,13], we find that knockdown of the lincRNAs studied here largely led to comparable numbers of activated and repressed genes (Supplementary Fig. 2 and Supplementary Table 4). To assess off-target effects further, we also profiled the effects of the second-best validated shRNA targeting 10 randomly selected lincRNA genes. In all cases, second shRNAs against the same target produced significantly similar expression changes (see Methods and Supplementary Table 5). These results indicate that the vast majority of lincRNAs have functional consequences on overall gene expression of comparable magnitude (in terms of number of affected genes and impact on levels) to the known transcriptional regulators in ES cells.

lincRNAs affect gene expression in *trans*

Following the observation that a few lincRNAs act in *cis*[24,25], some recent papers have claimed that most lincRNAs act primarily in *cis*[8,11,26]. We found no evidence to support this latter notion: knockdown of only 2 lincRNAs showed effects on a neighbouring gene, only 13 showed effects within a window of 10 genes on either side, and only 8 showed effects on genes within 300 kb; these proportions are no greater than observed for protein-coding genes (Supplementary Fig. 3 and Supplementary Table 6). In short, lincRNAs seem to affect expression largely in *trans*.

Our results contrast with a recent study that concluded that lincRNAs act in *cis*, based on the observation that knockdown of 7

out of 12 lincRNAs affected expression of a gene within 300 kb[11]. The explanation seems to be that the threshold in the previous study failed to account for multiple hypothesis testing within the local region. Accounting for this, the effects on neighbouring genes are no greater than expected by chance and are consistent with our observations here (see Methods).

Although some lincRNAs can regulate gene expression in *cis*[11,24,25], determining the precise proportion of *cis* regulators requires more direct experimental approaches. We note that our results are consistent with observed correlations between lincRNAs and neighbouring genes[2,26], which may represent shared upstream regulation[2,12] or local transcriptional effects[10,27]. In addition, the lincRNAs studied here should be distinguished from transcripts that are produced at enhancer sites[8,9], the function of which has yet to be determined.

lincRNAs maintain the pluripotent state

We next sought to investigate whether lincRNAs have a role in regulating the ES cell state. Regulation of the ES cell state involves two components: maintaining the pluripotency program and repressing differentiation programs[15]. To determine whether lincRNAs have a role in the maintenance of the pluripotency program, we studied their effects on the expression of Nanog, a key transcription factor that is required to establish[28] and uniquely marks the pluripotent state[29,30]. We infected ES cells carrying a luciferase reporter gene expressed from the endogenous *Nanog* promoter[31] with shRNAs targeting lincRNAs or protein-coding genes. We monitored loss of reporter activity after 8 days relative to 25 negative control hairpins across biological replicates (see Methods). To ensure that the observed effects were not simply due to a reduction in cell viability, we excluded shRNAs that caused a reduction in cell numbers (see Methods, Supplementary Fig. 4 and Supplementary Table 7). Altogether, we identified 26 lincRNAs that had major effects on endogenous Nanog levels with many at comparable levels to the knockdown of the known protein-coding regulators of pluripotency such as Oct4 and Nanog (Fig. 2a and Supplementary Table 7). This establishes that these lincRNAs have a role in maintaining the pluripotent state.

To validate further the role of these 26 lincRNAs in regulating the pluripotent state, we knocked down these lincRNAs in wild-type ES cells and measured mRNA levels of pluripotency marker genes *Oct4* (also called *Pou5f1*), *Sox2*, *Nanog*, *Klf4* and *Zfp42* after 8 days. In all cases we observed a significant reduction in the expression of multiple pluripotency markers with >90% showing a significant decrease in both *Oct4* and *Nanog* levels (Supplementary Fig. 5 and Supplementary Tables 8 and 9). To control for off-target effects, we studied additional hairpins targeting these lincRNAs. For 15 lincRNAs we had an effective second hairpin. In all 15 cases, the second hairpin produced comparable reductions in *Oct4* expression levels, showing that the observations were not due to off-target effects (Fig. 2b and Supplementary Table 10). Notably, >90% of lincRNA knockdowns affecting Nanog reporter levels led to loss of ES cell morphology (Fig. 2c and Supplementary Figs 6 and 7). Thus, inhibition of these 26 lincRNAs lead to an increased exit from the pluripotent state.

lincRNAs repress lineage programs

To determine if lincRNAs act in repressing differentiation programs we compared the overall gene expression patterns resulting from knockdown of the lincRNAs to published gene expression patterns resulting from induced differentiation of ES cells[32,33] and assessed significance using a permutation-derived FDR[34] (see Methods). These states include differentiation into endoderm, ectoderm, neuroectoderm, mesoderm and trophectoderm lineages. As a positive control for our analytical method, we confirmed the expected results that the expression pattern caused by Oct4 knockdown was strongly associated with the trophoectoderm lineage[35] and the pattern caused by Nanog knockdown was strongly associated with endoderm differentiation[30] (Fig. 3a).
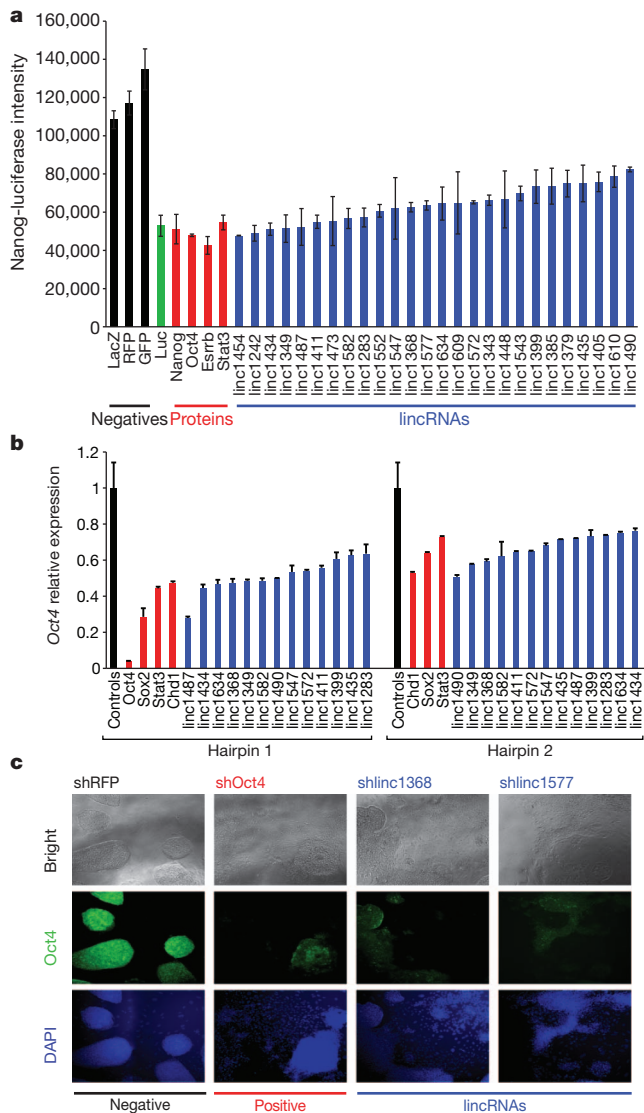
Figure 2 | lincRNAs are critical for the maintenance of pluripotency.
a, Activity from a *Nanog* promoter driving luciferase, following treatment with control hairpins (black) or hairpins targeting luciferase (green), selected protein-coding regulators (red), and lincRNAs (blue). **b**, Relative mRNA expression levels of *Oct4* after knockdown of selected protein-coding (red) and lincRNA (blue) genes affecting Nanog-luciferase levels. The best hairpin (Hairpin 1) and second best hairpin (Hairpin 2) are shown. All knockdowns are significant with a *P*-value <0.01. Error bars represent standard error (*n* = 4). **c**, Morphology of ES cells and immunofluorescence staining of Oct4 for a negative control hairpin (black line) and hairpins targeting Oct4 (red line), and two lincRNAs (blue line). The first row shows bright-field images, the second row shows immunofluorescence staining of the Oct4 protein, and the third row shows DAPI staining of the nuclei.

Using this approach, we identified 30 lincRNAs for which knockdown produced expression patterns similar to differentiation into specific lineages (Supplementary Table 11). Among these lincRNAs, 13 are associated with endoderm differentiation, 7 with ectoderm differentiation, 5 with neuroectoderm differentiation, 7 with mesoderm differentiation and 2 with the trophectoderm lineage (Fig. 3a). Consistent with these functional assignments, we observed that most (>85%) of the 30 lincRNAs associated with specific differentiation lineages showed upregulation of the well-known marker genes for the identified states[17,32] upon knockdown (such as *Sox17* (endoderm), *Fgf5* (ectoderm), *Pax6* (neuroectoderm), brachyury (mesoderm) and *Cdx2* (trophectoderm)) (Fig. 3b, Supplementary Figs 8 and 9 and Supplementary Tables 12 and 13).
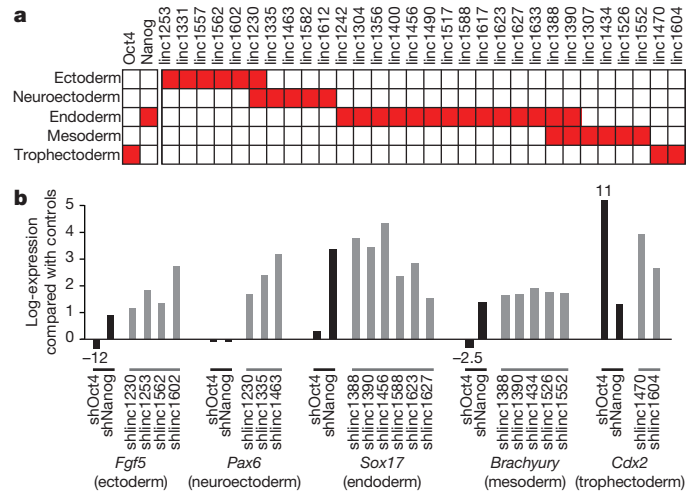
Figure 3 | lincRNAs repress specific differentiation lineages. a, Expression changes for each lincRNA compared to gene expression of five differentiation patterns. Each box shows significant positive association (red, FDR <0.01) for Oct4 and Nanog (left) and for lincRNAs (right). **b**, Expression changes upon knockdown of Oct4 and Nanog (black bars) and representative lincRNAs (grey bars) for five lineage marker genes. The expression changes (FDR <0.05) are displayed on a log scale as the *t*-statistic compared to a panel of negative control hairpins.

The fact that knockdown of these 30 lincRNAs induces gene expression programs associated with specific early differentiation lineages indicates that these lincRNAs normally are a barrier to such differentiation. Interestingly, most of the lincRNA knockdowns (~85%) that induce gene expression patterns associated with these lineages did not cause the cells to differentiate as determined by Nanog reporter levels (Supplementary Table 7) and Oct4 expression (Supplementary Fig. 10). This is consistent with observations for several critical ES cell chromatin regulators, such as the polycomb complex; loss-of-function of these regulators similarly induces lineage-specific markers without causing differentiation[18,36,37].

Together, these data indicate that many lincRNAs have important roles in regulating the ES cell state, including maintaining the pluripotent state and repressing specific differentiation lineages.

lincRNAs are targets of ES cell transcription factors

Having demonstrated a functional role for lincRNAs in ES cells, we sought to integrate the lincRNAs into the molecular circuitry controlling the pluripotent state. First, we explored how lincRNA expression is regulated in ES cells. Towards this end, we used published genome-wide maps of 9 pluripotency-associated transcription factors[16,38] and determined whether they bind to the promoters of lincRNA genes. Of the 226 lincRNA promoters ~75% are bound by at least 1 of 9 pluripotency-associated transcription factors (including Oct4, Sox2, Nanog, c-Myc, n-Myc, Klf4, Zfx, Smad and Tcf3) with a median of 3 factors bound to each promoter (Fig. 4a, Supplementary Fig. 11 and Supplementary Table 14), comparable to the proportion reported for protein-coding genes[16]. Interestingly, the three core factors (Oct4, Sox2 and Nanog) bind to the promoters of ~12% of all ES cell lincRNAs and ~50% of lincRNAs involved in the regulation of the pluripotent state.

To determine if lincRNA expression is functionally regulated by the pluripotency-associated transcription factors, we used shRNAs to knockdown the expression of 5 of the 9 pluripotency-associated transcription factor genes for which we could obtain validated hairpins and profiled the resulting changes in lincRNA expression after 4 days. Upon knockdown of a transcription factor, ~50% of lincRNA genes whose promoters are bound by the transcription factor exhibit expression changes (Fig. 4a); this proportion is comparable to that
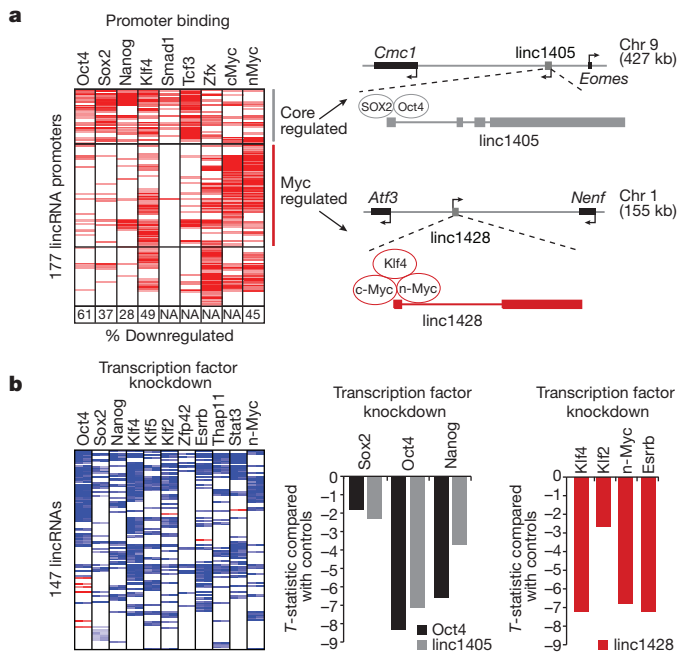
Figure 4 | lincRNAs are direct regulatory targets of the ES cell transcriptional circuitry. a, A heat-map representing ChIP-Seq enrichments for nine transcription factors (columns) at lincRNA promoters (rows). The percentage of bound lincRNAs downregulated upon knockdown of the transcription factor is indicated in boxes. NA, not measured. Right: examples of lincRNAs from two clusters ('core regulated' and 'Myc regulated') showing their genomic neighbourhood and transcription factor binding. **b**, Left: a heat-map representing changes in lincRNA expression (rows) after knockdown of 11 transcription factors (columns). Middle: effect of knockdown of Sox2, Oct4 and Nanog on expression levels of linc1405 (grey) and Oct4 (black). Right: effect of knockdown of Klf2, Klf4, n-Myc and Esrrb on expression levels of linc1428.

seen for protein-coding genes whose promoters are bound by the transcription factor (Supplementary Fig. 12). The strong but imperfect correlation between transcription-factor-binding and effect of transcription-factor knockdown is consistent with previous observations[39] and may reflect regulatory redundancy in the pluripotency network[40]. In addition, we profiled the knockdown of an additional 7 pluripotency-associated transcription factors (including Esrrb, Zfp42 and Stat3). Altogether, for ~60% of the ES cell lincRNAs, we identified a significant downregulation upon knockdown of 1 of these 11 transcription factors (Fig. 4b and Supplementary Table 15).

After retinoic-acid-induced differentiation of ES cells, the ES cell lincRNAs show temporal changes across the time course with ~75% showing a decrease in expression compared to untreated ES cells (Supplementary Fig. 13 and Supplementary Table 16). Notably, all of the lincRNAs shown to regulate pluripotency are downregulated upon retinoic acid treatment (Supplementary Fig. 13). Our results establish that lincRNAs are direct transcriptional targets of pluripotency-associated transcription factors and are dynamically expressed across differentiation. Collectively, these results demonstrate that lincRNAs are an important regulatory component within the ES cell circuitry.

lincRNAs bind diverse chromatin proteins

To explore how lincRNAs carry out their regulatory roles, we studied whether lincRNAs physically associate with chromatin regulatory proteins in ES cells. We previously showed that many human lincRNAs can interact with the polycomb repressive complex[4], a complex that has a critical functional role in the regulation of ES cells[18,19]. To determine whether the ES cell lincRNAs physically associate with the polycomb complex, we crosslinked RNA–protein complexes using formaldehyde, immunoprecipitated the complex using antibodies specific to both the Suz12 and Ezh2 components of polycomb, and

profiled the co-precipitated lincRNAs using a direct RNA quantification method[41] (see Methods). We performed immunoprecipitation of the polycomb complex across five biological replicates and eight mock-IgG controls, and we assessed significance using a permutation test (see Methods and Supplementary Fig. 16). Altogether, we identified 24 lincRNAs (~10% of the ES cell lincRNAs) that were strongly enriched for both polycomb components (Fig. 5b and Supplementary Table 17).

To determine if lincRNAs interact with additional chromatin proteins, we systematically analysed chromatin-modifying proteins that have been shown to have critical roles in ES cells[18–21,42]. Specifically, we screened antibodies against 28 chromatin complexes (see Methods, Supplementary Fig. 14 and Supplementary Table 18) and identified 11 additional chromatin complexes that are strongly and reproducibly associated with lincRNAs (see Methods and Supplementary Figs 15 and 16). These chromatin complexes are involved in 'reading' (Prc1, Cbx1 and Cbx3), 'writing' (Tip60/P400, Prc2, Setd8, Eset and Suv39h1) and 'erasing' (Jarid1b, Jarid1c, and Hdac1) histone modifications, as well as a chromatin-associated DNA binding protein (Yy1) (Fig. 5a). Altogether, we found that 74 (~30%) of the ES cell lincRNAs are associated with at least 1 of these 12 chromatin complexes (Fig. 5b and Supplementary Table 17). Although most of the identified interactions are with repressive chromatin regulators, this is probably due to limitations of our selection criteria and available antibodies.

Many lincRNAs are strongly associated with multiple chromatin complexes (Fig. 5b). For example, we identified 8 lincRNAs that bind to the Prc2 H3K27 and Eset H3K9 methyltransferase complexes (writers of repressive marks) and the Jarid1c H3K4 demethylase complex (an eraser of activating marks). Consistent with this, the Prc2 and Eset complexes have been reported to bind at many of the same 'bivalent' domains[21] and to associate functionally with the Jarid1c complex[43]. Similarly, we identified a distinct set of 17 lincRNAs that bind to the Prc2 complex ('writer' of K27 repressive marks), Prc1 complex ('reader' of K27 repressive marks) and Jarid1b complex ('eraser' of K4 activating
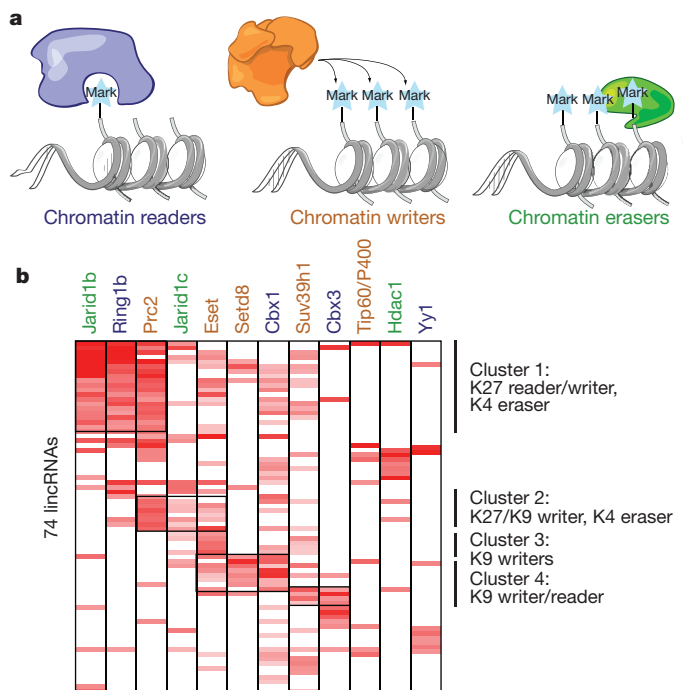
Figure 5 | lincRNAs physically interact with chromatin regulatory proteins. a, A schematic of the classes of chromatin regulators profiled: readers (blue), writers (orange) and erasers (green). **b**, A heat-map showing the enrichment of 74 lincRNAs (rows) for 1 of 12 chromatin regulatory complexes (columns). The names are colour-coded by chromatin-regulatory mechanism. Major clusters are indicated by vertical lines with a description of the chromatin components.

marks) (Fig. 5b), as well as other functionally consistent reader, writer and eraser combinations (Supplementary Fig. 17). One of several potential models consistent with these data are that lincRNAs may bind to multiple distinct protein complexes, perhaps serving as 'flexible scaffolds' to bridge functionally related complexes as previously described for telomerase RNA[44].

To determine if the identified lincRNA–protein interactions have a functional role, we examined the effects on gene expression resulting from knockdown of individual lincRNAs that are physically associated with particular chromatin complexes and from knockdown of genes encoding the associated complex itself (see Methods). For >40% of these lincRNA–protein interactions, we identified a highly significant overlap in affected gene expression programs compared to just ~6% for random lincRNA–protein pairs (see Methods and Supplementary Table 19). Other cases may reflect the limited power to detect the overlaps, because specific lincRNA–protein complexes may be related to only a fraction of the overall expression pattern mediated by the chromatin complex.

Together, these data demonstrate that many ES cell lincRNAs physically associate with multiple different chromatin regulatory proteins and that these interactions are probably important for the regulation of gene expression programs.

Discussion

Although the mammalian genome encodes thousands of lincRNA genes, few have been functionally characterized. We performed an unbiased loss-of-function analysis of lincRNAs expressed in ES cells and show that lincRNAs are clearly functional and primarily act in *trans* to affect global gene expression. We establish that lincRNAs are key components of the ES cell transcriptional network that are functionally important for maintaining the pluripotent state, and that many are downregulated upon differentiation. The ES cell lincRNAs physically interact with chromatin proteins, many of which have been previously implicated in the maintenance of the pluripotent state[18,20,21]. In addition to chromatin proteins, lincRNAs interact with other protein complexes including many RNA-binding proteins (data not shown).

Our data suggest a model whereby a distinct set of lincRNAs is transcribed in a given cell type and interacts with ubiquitous regulatory protein complexes to form cell-type-specific RNA–protein complexes that coordinate cell-type-specific gene expression programs (Fig. 6). Because many of the lincRNAs studied here interact with multiple different protein complexes, they may act as cell-type-specific 'flexible scaffolds'[44] to bring together protein complexes into larger functional units (Fig. 6). This model has been previously demonstrated for the yeast telomerase RNA[44] and suggested for the XIST[45] and HOTAIR[46] lincRNAs. The hypothesis that lincRNAs serve as flexible scaffolds could explain the uneven patterns of evolutionary conservation seen across the length of lincRNA genes[3]: the more highly conserved patches could correspond to regions of interaction with protein complexes.

Although a model of lincRNAs acting as 'flexible scaffolds' is attractive, it is far from proven. Testing the hypothesis for lincRNAs will require systematic studies, including defining all protein complexes with which lincRNAs interact, determining where these protein interactions assemble on RNA, and ascertaining whether they bind simultaneously or alternatively. Moreover, understanding how lincRNA–protein interactions give rise to specific patterns of gene expression will require determination of the functional contribution of each interaction and possible localization of the complex to its genomic targets.

METHODS SUMMARY

RNAi expression effects. We cloned five shRNAs targeting each lincRNA into a puromycin-resistant lentiviral vector[22]. ES cells were plated on pre-gelatinized 96-well plates and infected with lentivirus before addition of irradiated DR4 mouse embryonic fibroblasts (MEFs). Media containing $1\,\mu g\,ml^{-1}$ puromycin was added 24 h after infection. On-target knockdown was assessed after 4 days and the best hairpin showing a knockdown >60% was selected. RNA from 147
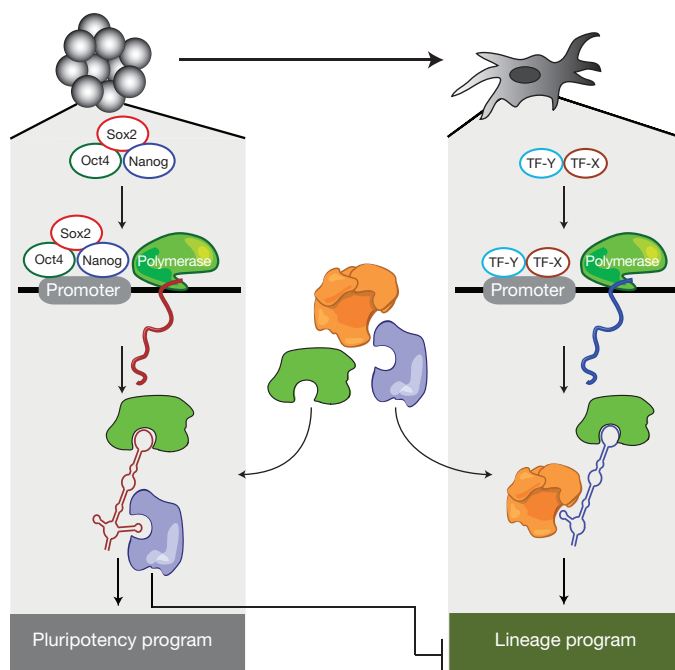
Figure 6 | A model for lincRNA integration into the molecular circuitry of the cell. ES-cell-specific transcription factors (such as Oct4, Sox2 and Nanog) bind to the promoter of a lincRNA gene and drive its transcription. The lincRNA binds to ubiquitous regulatory proteins, giving rise to cell-type-specific RNA–protein complexes. Through different combinations of protein interactions, the lincRNA–protein complex can give rise to unique transcriptional programs. Right: a similar process may also work in other cell types with specific transcription factors regulating lincRNAs, creating cell-type-specific RNA–protein complexes and regulating cell-type-specific expression programs.

lincRNAs, 40 protein-coding genes and 27 negative controls were hybridized to Agilent microarrays. Differentially expressed genes were defined as having an FDR <5% and fold-change >2-fold compared to controls.

Screening for pluripotency effects. Nanog-luciferase ES cells[31] were infected and measured after 8 days. Hits were identified if they reduced luciferase levels ($z < -6$) across all replicates and did not reduce AlamarBlue levels. Hits were validated in wild-type ES cells by measuring mRNA levels of *Oct4*, *Nanog*, *Sox2*, *Klf4* and *Zfp42*. Oct4 expression was assessed using immunofluorescence staining and morphology was visually assessed.

Lineage expression effects. Lineage expression programs were defined using published data sets (Gene Expression Omnibus GSE12982, GSE11523, and GSE4082) and curated gene expression signatures[32,33]. Overlaps in gene expression effects were assessed using a modified GSEA[34]. Expression changes in lineage markers were determined using qPCR.

Transcription factor binding and regulation. ChIP-Seq data was downloaded (GSE11724 and GSE11431), aligned and analysed. lincRNA promoters were previously defined using H3K4me3 peaks[3]. Changes in expression of the lincRNAs upon knockdown of the transcription factors were analysed using Agilent microarrays.

Chromatin binding and overlap in expression. ES cells were crosslinked with formaldehyde, lysed, immunoprecipitated, washed and reverse crosslinked. RNA was hybridized to the Nanostring code set. We tested antibodies for 28 chromatin complexes and selected successful antibodies that had >10 lincRNAs exceeding a fivefold change and had significant enrichments across 3 replicates. We compared the overlap in gene expression using a modified GSEA[34].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. The FANTOM Consortium. The transcriptional landscape of the mammalian genome. *Science* **309,** 1559–1563 (2005).
2. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458,** 223–227 (2009).

3. Guttman, M. et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nature Biotechnol. **28,** 503–510 (2010).
4. Khalil, A. M. et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl Acad. Sci. USA **106,** 11667–11672 (2009).
5. Ponjavic, J., Ponting, C. P. & Lunter, G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res. **17,** 556–565 (2007).
6. Mattick, J. S. The genetic signatures of noncoding RNAs. PLoS Genet. **5,** e1000459 (2009).
7. Koziol, M. J. & Rinn, J. L. RNA traffic control of chromatin complexes. Curr. Opin. Genet. Dev. **20,** 142–148 (2010).
8. De Santa, F. et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol. **8,** e1000384 (2010).
9. Kim, T. K. et al. Widespread transcription at neuronal activity-regulated enhancers. Nature **465,** 182–187 (2010).
10. Ebisuya, M., Yamamoto, T., Nakajima, M. & Nishida, E. Ripples from neighbouring transcription. Nature Cell Biol. **10,** 1106–1113 (2008).
11. Ørom, U. A. et al. Long noncoding RNAs with enhancer-like function in human cells. Cell **143,** 46–58 (2010).
12. Huarte, M. et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell **142,** 409–419 (2010).
13. Rinn, J. L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell **129,** 1311–1323 (2007).
14. Smith, A. G. Embryo-derived stem cells: of mice and men. Annu. Rev. Cell Dev. Biol. **17,** 435–462 (2001).
15. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell **132,** 567–582 (2008).
16. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell **133,** 1106–1117 (2008).
17. Ivanova, N. et al. Dissecting self-renewal in stem cells with RNA interference. Nature **442,** 533–538 (2006).
18. Boyer, L. A. et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature **441,** 349–353 (2006).
19. Bernstein, B. E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell **125,** 315–326 (2006).
20. Fazzio, T. G., Huff, J. T. & Panning, B. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. Cell **134,** 162–174 (2008).
21. Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. & Young, R. A. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. Genes Dev. **23,** 2484–2489 (2009).
22. Moffat, J. et al. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell **124,** 1283–1298 (2006).
23. Hu, G. et al. A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes Dev. **23,** 837–848 (2009).
24. Plath, K., Mlynarczyk-Evans, S., Nusinow, D. A. & Panning, B. Xist RNA and the mechanism of X chromosome inactivation. Annu. Rev. Genet. **36,** 233–278 (2002).
25. Koerner, M. V., Pauler, F. M., Huang, R. & Barlow, D. P. The function of non-coding RNAs in genomic imprinting. Development **136,** 1771–1783 (2009).
26. Ponjavic, J., Oliver, P. L., Lunter, G. & Ponting, C. P. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. PLoS Genet. **5,** e1000617 (2009).
27. Sproul, D., Gilbert, N. & Bickmore, W. A. The role of chromatin structure in regulating the expression of clustered genes. Nature Rev. Genet. **6,** 775–781 (2005).
28. Silva, J. et al. Nanog is the gateway to the pluripotent ground state. Cell **138,** 722–737 (2009).
29. Chambers, I. et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell **113,** 643–655 (2003).
30. Mitsui, K. et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell **113,** 631–642 (2003).
31. Brambrink, T. et al. Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell **2,** 151–159 (2008).
32. Sherwood, R. I. et al. Prospective isolation and global gene expression analysis of definitive and visceral endoderm. Dev. Biol. **304,** 541–555 (2007).
33. Aiba, K. et al. Defining developmental potency and cell lineage trajectories by expression profiling of differentiating mouse embryonic stem cells. DNA Res. **16,** 73–80 (2009).
34. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl Acad. Sci. USA **102,** 15545–15550 (2005).
35. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nature Genet. **24,** 372–376 (2000).
36. Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M. & Helin, K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Mol. Cell. Biol. **27,** 3769–3779 (2007).
37. Jiang, H. et al. Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. Cell **144,** 513–525 (2011).
38. Marson, A. et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell **134,** 521–533 (2008).
39. Kunarso, G. et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nature Genet. **42,** 631–634 (2010).
40. Jiang, J. et al. A core Klf circuitry regulates self-renewal of embryonic stem cells. Nature Cell Biol. **10,** 353–360 (2008).
41. Geiss, G. K. et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nature Biotechnol. **26,** 317–325 (2008).
42. Dey, B. K. et al. The histone demethylase KDM5b/JARID1b plays a role in cell fate decisions by blocking terminal differentiation. Mol. Cell. Biol. **28,** 5312–5327 (2008).
43. Cloos, P. A., Christensen, J., Agger, K. & Helin, K. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes Dev. **22,** 1115–1140 (2008).
44. Zappulla, D. C. & Cech, T. R. Yeast telomerase RNA: a flexible scaffold for protein subunits. Proc. Natl Acad. Sci. USA **101,** 10024–10029 (2004).
45. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nature Genet. **30,** 167–174 (2002).
46. Tsai, M. C. et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science **329,** 689–693 (2010).

METHODS

ES cell culture. V6.5 (genotype 129SvJae × C57BL/6) and Nanog-luciferase[31] ES cells were co-cultured with irradiated C57BL/6 MEFs (GlobalStem; GSC-6002C) on pre-gelatinized plates as previously described[47]. Briefly, cells were cultured in mES media consisting of knockout DMEM (Invitrogen; 10829018) supplemented with 10% FBS (GlobalStem; GSM-6002), 1% penicillin-streptomycin (Invitrogen; 15140-163), 1% L-glutamine (Invitrogen; 25030-164), 0.001% β-mercaptoethanol (Sigma; M3148-100ML) and 0.01% ESGRO (Millipore; ESG1106).

Picking lincRNA gene candidates. Using our previous catalogue of K4-K36 defined lincRNAs[2] along with the reconstructed full-length sequences we determined using RNA-Seq[3], we designed shRNA hairpins targeting each lincRNA identified in both sets. Specifically, we used the conservative K4-K36 definitions from our previous work[2] that were expressed in mouse ES cells. We further filtered the list to include only multi-exonic lincRNAs that were reconstructed in mouse ES cells[3]. Together, this yielded 226 lincRNA genes.

Picking protein-coding gene candidates. We selected protein coding gene controls consisting of both transcription factors and chromatin proteins. These proteins were selected based on their well-characterized role in regulating mouse ES cells and include Oct4 (Pou5f1)[35,48], Sox2 (refs 17, 49) Nanog (refs 29, 30), Stat3 (ref. 50), Klf4 (ref. 51) and Zfp42 (Rex1)[52]. In addition, we selected additional transcriptional and chromatin regulators that were identified by RNAi screens as regulators of pluripotency[17,20,23] and/or were found in smaller focused studies to have critical roles in the maintenance of the pluripotent state (such as Carm1 (ref. 53), Chd1 (ref. 54), Thap11 (ref. 55), Suz12 (refs 18, 19, 36) and Setdb1 (refs 21, 56)). A full list is provided in Supplementary Table 2.

shRNA design rules. For each lincRNA we designed five hairpins by extending the previously described design rules[22] accounting for the sequence content of the hairpin, miRNA seed matches, uniqueness to the target compared to the transcriptome and the genome, and number of lincRNA isoforms covered.

For each lincRNA we enumerated all 21-mer sub-sequences and scored them as follows: (1) A 'clamp score' was computed by looking at the nucleotides at positions 18, 19 and 20. If all three positions contained an A/T it was assigned a score of 4, if two positions were A/T it was assigned a score of 1.5 and if one was A/T it was assigned a score of 0.8. We then looked at positions 16, 17, and 21; if all three were A/T it was assigned a score of 1.25, if two were A/T it was assigned a score of 1.1, and if one was A/T is assigned a score of 0.8. The clamp score was computed as the product of these two scores. (2) A 'GC score' was computed by looking at the total GC percentage of the 21-mer sequence. If the sequence was <25% GC it was assigned a score of 0.01, if it was <55% it was assigned a score of 3, if it was <60% it was assigned a score of 1, and if >60% it was assigned a score of 0.01. (3) A '4-mer penalty' of 0.01 was assigned for any hairpin containing the same nucleotide in 4 subsequent nucleotides. (4) A '7 GC penalty' of 0.01 was assigned to any hairpin containing any 7 consecutive G/C nucleotides. (5) We removed all hairpins containing an A in either position 1 or position 2 of the hairpin. (6) We removed all hairpins containing a repeat masked nucleotide. (7) Finally, we computed a 'miRNA-seed penalty' by looking at the forward positions 11–17, 12–20 and 13–19 of the hairpin as well as the reverse complement of positions 14–20, 15–21, or 16–21 plus a 3′ C. We then looked up whether these positions matched known miRNA seeds and with what frequency. We computed the scores for the forward and reverse positions and defined the score as the product of the forward and reverse scores. The final score for each hairpin sequence is defined as the product of all seven scores.

We then sorted the candidate hairpin sequences by score, breaking high-scoring ties by the total number of lincRNA isoforms that are covered by the hairpin. We then aligned each hairpin sequence against both the genome and the RefSeq-defined transcriptome (NCBI Release 39), and filtered any hairpin with fewer than three mismatches to any other gene or position in the genome. Candidate sequences were chosen for shRNA production by first picking the highest scoring candidate and then proceeding to successively lower scores. As each hairpin was selected, all other hairpins overlapping this hairpin were removed. We repeated this process until we identified five hairpins that covered each lincRNA.

shRNA cloning and virus prep. We designed 1,143 hairpins targeting 226 lincRNA genes. Of these, we successfully cloned 1,010 hairpins targeting 214 lincRNAs. These hairpins were cloned into a vector containing a puromycin resistance gene and incorporated into a lentiviral vector as previously described[22]. Briefly, synthetic double-stranded oligos that represent a stem-loop hairpin structure were cloned into the second-generation TRC (the RNAi Consortium) lentiviral vector, pLKO.5; the expression of a given hairpin produces a shRNA that targets the gene of interest. Lentivirus was prepared as previously described[22]. Briefly, 100 ng of shRNA plasmid, 100 ng of packaging plasmid (psPAX2) and 10 ng of envelope plasmid (VSV-G) were used to transfect packaging cells (293T) with TransIT-LT1 (Mirus Bio). Virus was harvested 48 and 70 h after

transfection. Two harvests were combined. Virus titres were measured as previously described[22]. Briefly, we measured virus titres by infecting A549 cells with appropriately diluted viruses. Twenty-four hours after infection, puromycin was added to a final concentration of $5\,\mu g\,ml^{-1}$ and the selection proceeded for 48 h. The number of surviving cells, which is correlated to virus titre, was measured by AlamarBlue (BioSource) staining using the Envision 2103 Multilabel plate reader (PerkinElmer).

Infection and selection protocol. V6.5 ES cells or Nanog-luciferase ES cells were plated at a density of 5,000 cells per well (8-day time point) or 25,000 cells per well (4-day time point) in $100\,\mu l$ mES media onto pre-gelatinized 96-well dishes (VWR; BD356689). Cells were infected with $5\,\mu l$ of a lentiviral shRNA stock and incubated at $37\,^{\circ}C$ for 30 min. Puromycin-resistant DR4 MEFs (GlobalStem; GSC-6004G) were then added to the plates at a density of ~6,000 cells per well and incubated overnight at $37\,^{\circ}C$, 5% CO_2. After 24 h, all media was removed from the cells and replaced with media containing $1\,\mu g\,ml^{-1}$ puromycin. Media was then changed every other day with fresh media containing $1\,\mu g\,ml^{-1}$ puromycin. The end-point depended on the assay and was either 4 days after infection (knockdown validation and microarrays) or 8 days (reporters and qPCR of marker genes).

RNA extraction. ES cells were infected and lysed at day 4 with $150\,\mu l$ of Qiagen's RLT buffer and three replicates of each virus plate were pooled for RNA extraction using Qiagen's RNeasy 96-well columns (74181). RNA extraction was completed following Qiagen's RNeasy 96-well protocol with the following modifications: $450\,\mu l$ of 70% ethanol was added to $450\,\mu l$ total lysate before the first spin. An additional RPE wash was added to the protocol, for a total of three RPE washes.

lincRNA primer design and pre-screen. lincRNA primers were designed using primer3 (http://frodo.wi.mit.edu/primer3/). Specifically, we designed primers spanning exon–exon junctions by specifying each of the regions as preferred inclusion regions in the primer3 program. When a low-scoring primer pair (primer penalty <1) was available it was used. If none was available, we then identified all primers that contained amplicons that spanned an exon–exon junction. In a few cases, when we could not identify a primer pair spanning an exon–exon junction, we designed primers within an exon of the lincRNA. For each primer pair, we tested the specificity against the transcriptome[57] (RefSeq NCBI Release 39) and the genome (Mouse MM9) using the isPCR (http://genome.ucsc.edu/cgi-bin/hgPcr) program. Specifically, we required that the primer pair amplify the lincRNA gene and no other genomic of gene amplicon.

For each primer pair, we validated the quantification and specificity before use. Specifically, we tested primers in qPCR reactions using a dilution series of mouse ES cDNA including a no reverse transcriptase (RT) sample. We excluded any primer that did not have robust quantification across a 64-fold dilution curve, had high signal in the no RT sample, or had low detectable expression in the undiluted sample (cycle number >34). For primers that failed this validation we redesigned and tested new primers.

Knockdown validation using qPCR. To determine if lincRNA hairpins were effective at knocking down the lincRNA of interest, we infected each hairpin into mouse embryonic stem cells, selected for lentiviral integration, and measured changes in the targeted lincRNA expression level. We isolated total cellular RNA after 4 days; this time point was chosen to allow for identification of robust changes while minimizing secondary effects due to differentiation of the ES cells. We reasoned that this would allow us to determine more direct effects due to RNAi rather than to differentiation.

Gene panels were constructed that contained all five hairpins targeting a gene along with an empty vector control pLKO.5-nullT and the GFP-targeting hairpin clonetechGfp_437s1c1. cDNA was generated using $10\,\mu l$ of RNA and $10\,\mu l$ of $2\times$ cDNA master mix containing $5\times$ Transcriptor RT Reaction Buffer (Roche), DTT, MMLV-RT (Roche), dNTPs (Agilent; 200415-51), Random 9-mer oligos (IDT), Oligo-dT (IDT) and water. cDNA was diluted 1:9 and quantitative PCR was performed using 250 nM of each primer in $2\times$ Sybr green master mix (Roche) and run on a Roche Light-Cycler 480. Target lincRNA expression and Gapdh levels were computed for each panel. lincRNA expression levels were normalized by Gapdh levels and this normalized value was compared to the reference control hairpins within the panel. Knockdown levels were computed as the average of the fold decrease compared to the two control hairpins. Hairpins showing a knockdown greater than 60% of the endogenous level were considered validated and the best validated hairpin from a lincRNA panel was selected for microarray studies.

Picking candidates for microarray analysis. To assess the effects of a lincRNA on gene expression, we profiled the changes in gene expression after knocking down each lincRNA gene. Specifically, for each lincRNA with at least one validated hairpin we profiled the genome-wide expression level changes after knockdown across two independent infections (see above). To control for expression

changes due to viral infection, we performed five independent infections containing no RNAi hairpin (pLKO.5-nullT). This control hairpin was embedded in each RNA preparation plate. To control for effects due to an off-target RNAi effect, we profiled 27 distinct negative control hairpins which do not have a known target in the cell. These hairpins included 6 RFP hairpins, 10 GFP hairpins, 6 luciferase hairpins and 5 LacZ hairpins. These hairpins provide a measurement of the variability of the RNAi response triggered due to nonspecific effects. Furthermore, we profiled hairpins targeting 147 lincRNAs, including 10 with a second best hairpin, and 40 protein-coding genes in biological replicate. The hairpins and their replicates were randomly distributed across 7 96 well plates and prepared in batches. Each RNA preparation batch contained one pLKO hairpin and one clonetechGfp_437s1c1 hairpin in a random location on the plate. To minimize batch effects, the plate locations of the biological replicates were scrambled and the positions within the plates were scrambled for all hairpins and replicates.

Agilent microarray hybridization. Using Agilent's One-Colour Quick Amp Labelling kit (5190-0442), we amplified and labelled total RNA for hybridization to prototype mouse lincRNA arrays (G4140-90040) according to manufacturer's instructions with a few variations. The custom Agilent SurePrint G3 8x60K mouse array design used for this study (G4102A, AMADID 025725 G4852A) has probes to 21,503 Entrez genes and 2,230 lincRNA genes. A new updated version of this mouse design is commercially available that contains probes to 34,017 Entrez gene targets as well as 2,230 lincRNA genes (G4825A). The cRNA samples were prepared by diluting 200 ng of RNA in 8.3 µl water and adding positive control one-colour RNA spike-in mix (Agilent, 5188-5282) that was diluted serially 1:20, then 1:25 and finally 1:10. We annealed the T7 promoter primer from the kit by incubating at 65 °C for 10 min. We prepared the cDNA master mix and added it to the annealed RNA and incubated at 40 °C for 2 h, followed by 65 °C for 15 min. We prepared the cRNA transcription master mix and added it to the cDNA and incubated at 40 °C for 2 h protected from light. We purified the labelled cRNA using Qiagen's RNeasy 96-well columns (Qiagen, 74181) by adding 350 µl of Qiagen RLT (without BME) to the cRNA followed by the addition of 250 µl of 95% ethanol before applying to the plate column. After a 4 min spin at 6,000 r.p.m., we washed the columns three times with 800 µl buffer RPE. We dried the columns by spinning for 10 min and eluted the cRNA with 50 µl of water. We measured the cRNA yield and dye incorporation using the Nanodrop 8000 Microarray measurement setting. We mixed 600 ng of cRNA with blocking agent and fragmentation buffer (Agilent, 5190-0404) and fragmented for 30 min in the dark at 60 °C. We added 2× hybridization buffer to each sample and loaded 40 µl onto an 8-pack Hybridization gasket. We placed the microarray slides on top, sealed in the hybridization chamber, and incubated for 18 h at 65 °C. We washed the slides for 1 min in room temperature GE Wash Buffer 1 and then for 1 min in 37 °C GE Wash Buffer 2 (Agilent 5188-5327, no triton addition). We scanned the microarrays using an Agilent Scanner C (G2565CA) using the following settings: dye channel = red & green, scan region = scan area (61 × 21.6 mm), scan resolution = 3 µm. We prepared all of the samples simultaneously using homogenous master mixes to limit variability. Fragmentation and hybridization was staggered over time in batches of 3 to 4 slides (24 to 32 samples).

Array filtering, normalization and probe filtering. Each array was processed and data extracted using the Agilent feature extraction software (G4462AA, Version 10.7.3). Samples were retained if they passed all the following quality control statistics: AnyColourPrcntFeatNonUnifOL <1; eQCOneColourSpikeDetectionLimit >0.01 and <2.0; Metric_absGE1E1aSlope between 0.9 and 1.2; Metric_gE1aMedCV ProcSignal <8; gNegCtrlAveBGSubSig >−10 and <5; Metric_gNegCtrlAveNet Sig <40; gNegCtrlSDevBGSubSig <10; Metric_gNonCntrlMedCVProcSignal <8; Metric_gSpatialDetrendRMSFilterMinusFit <15; SpotAnalysis_PixelSkew CookiePct >0.8 and <1.2.

Gene expression values were determined using the gProcessedSignal intensity values. Probes were flagged if they were not detectable well above background or had an expression level lower than the lowest detectable spike-in control value. The values were floored across all samples by taking the maximum of the minimum non-flagged values across all experiments. Any value less than this maximum value was set to the maximum. This conservatively eliminates any detection variability across the samples due to stringency or other array variables.

The result of this is a single value for each probe per array. To normalize expression values across arrays, we performed quantile normalization as previously described[58]. Briefly, we ranked each array from lowest to highest expression. For each rank, we computed the average expression and each experiment with this value at the associated rank. For each probe, we computed the difference between the second smallest expression value and the second largest expression value. If this difference was less than 2, we filtered the probe. This metric was chosen to eliminate bias due to single sample outliers.

Identifying significant gene expression hits from RNAi knockdowns. To control for effects due to nonspecific effects of shRNAs, we profiled 27 distinct

negative control hairpins which do not have a known target in the cell. These hairpins provide a measurement of the variability of the expression profiles due to random variability or triggered by 'off-target' effects of the shRNA lentiviruses. Assuming that any observed effects in the negative control hairpins are due to off-target effects and observed effects in the targeting hairpins include a mix of both off-target effects and on-target effects, we use permutations of the negative controls to assign a FDR confidence level for being an on-target hit to each gene. As such, a gene would only reach genome-wide significance if the number of genes and scale of the effect was much larger than would be observed randomly among all of the expression changes found for the negative control hairpin.

Specifically, for each gene we computed a t-statistic between shRNAs targeting the lincRNA and control shRNA samples. To assess the significance of each gene we permuted the sample and control groups retaining the relative sizes of the groups and computing the same t-statistic. We then assigned an FDR value to each gene by computing the average number of values in the permuted t-statistics that were greater than the observed value of interest and divided this by the number of all observed t-statistics that were greater than the observed value. We defined genes as significantly differentially expressed if the FDR was <5% and the fold-change compared to the negative controls was >2-fold. Using this approach, an effect would only reach a significant FDR if the scale is significantly larger than would be observed in the negative controls. Knockdown of a lincRNA was considered to have a significant effect on gene expression if we identified at least 10 genes that had an effect that passed all of the criteria.

Gene-neighbour analysis. We identified neighbouring genes based on the RefSeq genome annotation[57] (NCBI Release 39). We excluded from analysis all RefSeq genes that corresponded to our lincRNA of interest but included all other coding and non-coding transcripts. We identified a significant hit as any lincRNA affecting a neighbour within 10 genes on either side with an FDR<0.05 and twofold expression change. To compute the closest affected neighbour, we classified all genes affected upon knockdown of the lincRNAs using the same criteria above. We computed the distance between each affected gene and the locus of the lincRNA gene (and protein-coding gene) that was perturbed and took the minimum absolute distance across all affected genes.

Analysis of expected number of neighbouring genes that will change by chance. To determine the expected number of differentially expressed 'neighbouring' genes occurring by chance assuming that the knockdown has no effect on gene expression, we calculated the average number of genes in a 300-kb window around a randomly selected gene in the human and mouse genome. We calculated this to be 11.2 (human) and 11.8 (mouse). For simplicity, we will conservatively round this down to 11. Assuming that no genes are changing between the knockdown and control, using a nominal P-value, which has a uniform distribution under the null hypothesis (nothing effected), we would expect to see a difference called in 5% of cases at a P-value of 0.05. If we test one locus, which has on average 11 neighbours, we would expect to identify 0.55 hits by chance (11 × 0.05 = 0.55). However, if we now test 12 loci we would expect to see 6.6 (12 × 0.55) knockdowns that appear to have an effect under the null hypothesis.

Luciferase analysis of Nanog ES lines. ES cells containing a Nanog-luciferase construct[31] were infected in biological duplicate and monitored after 7 days. Luciferase activity was measured using Bright-Glo (Promega). All reagents and cells were equilibrated to room temperature. 100 µl Bright-Glo solution was added to each plate well. Plates were incubated in the dark at room temperature for 10 min and luciferase was measured on a plate reader. The luciferase units were normalized to the control hairpins and a Z-score compared to the negative controls (excluding luciferase hairpins) was computed. For each hairpin, we computed a Z-score relative to the negative control hairpins and identified hits reducing luciferase levels more than 6 standard deviations ($Z < -6$) for both independent replicates. In all cases we were able to identify a significant reduction in luciferase levels when using distinct hairpins targeting luciferase. To exclude hits that were due to an overall reduction in proliferation (which would also cause a reduction of Nanog positive cells in this read-out) we excluded all hairpins that caused a reduction in proliferation as measured by AlamarBlue incorporation (described below). AlamarBlue incorporation was measured in the same cells immediately before reading out Nanog-luciferase levels.

AlamarBlue analysis of ES lines. After a 7-day infection, Nanog-luciferase cell viability was measured using AlamarBlue (Invitrogen; DAL1025). AlamarBlue was mixed with mES media in a 1:10 ratio, added to the cells and incubated at 37 °C for 1 h. Absorbance readings at 570 nm were taken. To control for possible effects due to virus titre, we measured AlamarBlue incorporation on both puromycin treated and non-puromycin treated samples for each infection.

mRNA analysis of pluripotency markers. V6.5 ES cells were infected with shRNAs targeting lincRNAs, protein-coding genes, and 21 negative controls. After 8 days, RNA was extracted and mRNA levels of the *Oct4*, *Nanog*, *Sox2*,

Klf4 and *Zfp42* pluripotency markers were analysed using qPCR. Primer sequences are listed in Supplementary Table 9. Each sample was normalized to *Gapdh* levels. Significance was assessed compared to the negative control hairpins using a one-tailed *t*-test.

To control for off-target effects, we analysed additional hairpins against the 26 lincRNAs affecting Nanog-luciferase levels. Of the 26 lincRNAs, we identified 15 lincRNAs that contained an additional hairpin that reduced lincRNA expression by >50%. V6.5 ES cells were infected with the best and additional hairpin across biological replicates for these 15 lincRNAs and 21 negative control hairpins. RNA was extracted after 8 days and Oct4 expression levels were determined using qPCR. Significance was assessed relative to the negative controls using a one-tailed *t*-test.

Immunofluorescence. We crosslinked cells in 4% paraformaldehyde for 15 min, and washed in 1× PBS three times. To permeabilize the cells, we washed with 1× PBS +0.1% Triton and then blocked in 1× PBS + 0.1% Triton + 1% BSA for 45 min at room temperature. We incubated cells with anti-Pou5f1 antibody (Santa Cruz: SC-9081) at 1:100 dilution in blocking solution for 1.5 h at room temperature and then washed in blocking solution three times. Next, we incubated cells in anti-rabbit secondary antibody coupled to GFP (Jackson ImmunoResearch: 111-486-152) at a dilution of 1:1,000 in blocking solution for 45 min. Finally, we thoroughly washed cells in blocking solution three times, and added vectashield containing DAPI (VWR: 101098-044) to each well.

Public data set curation. Traditionally, lineage markers are used to identify changes in phenotypic states. Although these markers can be good indicators of differentiation potential, there are two major limitations with this approach. First, there are multiple genes that are associated with each lineage so simply looking at one can often be misleading. Second, this approach only works for classifying states with well-characterized marker genes but would not work for a comprehensive characterization of the function in the cell. Therefore, we decided to take a different approach and look at the entire gene expression profile of each lincRNA knockdown to determine what cell state each lincRNA resembles.

We curated a set of ES perturbations and differentiation states from publicly available sources. Specifically, we used the NCBI e-utils (http://eutils.ncbi.nlm.nih.gov/) to programmatically identify all published data sets containing keywords associated with embryonic stem cells. We filtered the list to only include mouse data sets that were generated across one of three commercial array platforms (Affymetrix, Agilent and Illumina). Following this approach, we manually curated the list to include data sets associated with ES cell perturbations (genetic deletions, RNAi, or chemical perturbations) and differentiation or induced differentiation profiles. This curation yielded 41 GEO data sets corresponding to >150 samples.

Specifically, we defined differentiation lineage states using the following data sets. (1) Neuroectoderm: we downloaded a data set (GSE12982) corresponding to mouse ES cells containing a Sox1–GFP reporter construct. Upon differentiation of Sox1–GFP ES cells into embryoid bodies (EBs), Sox1–GFP-positive cells were collected and their global expression was profiled[59]. In addition, we downloaded a data set (GSE4082)[60] corresponding to direct neuroectoderm differentiation[61].

(2) Mesoderm: we downloaded the same data set (GSE12982) as above, where the authors differentiated brachyury–GFP reporter ES cells into EBs and sorted and profiled brachyury–GFP-positive cells[59].

(3) Endoderm: we downloaded a data set (GSE11523) corresponding to mouse ES cells which were engineered to overexpress GATA6[33]. GATA6 overexpression has been shown to drive ES cells into a primitive endoderm-like state[62].

(4) Ectoderm: we downloaded a data set (GSE4082)[60] corresponding to mouse ES cells differentiated into primitive ectoderm-like cells with defined media[61].

(5) Trophectoderm: we downloaded a data set (GSE11523)[33] corresponding to mouse ES cells which were engineered to deplete Oct4[35]. These cells have been shown to enter a trophectoderm-like state[35]. To ensure specificity to the trophectoderm state, we also compared the expression effects to trophoblast stem cells[33]. For all lincRNAs identified, we required a significant enrichment for both induced Oct4 knockout and trophoblast stem cell programs.

In addition, for all lineage states we used a curated discrete gene expression signature of differentiation which was previously functionally tested and shown to correspond specifically to differentiation into the associated states[63].

Continuous enrichment analysis and phenotype-projection analysis. To determine relationships between lincRNA knockdowns and functional states, we used a modified Gene Set Enrichment Analysis[34] approach that accounts for the continuous nature of the two data sets, similar to previously described extensions[34,64,65]. For each lincRNA knockdown by functional pair we compute a continuous enrichment score. Specifically, (1) for each lincRNA knockdown we compute a normalized score matrix compared to a panel of negative control hairpins by computing a *t*-statistic for each gene between the replicate lincRNA knockdown expression values and the control knockdown values. (2) For each experiment, we sort the matrix by the normalized score such that the most

differentially expressed upregulated gene is first and the most differentially expressed downregulated gene is last. Using this ordering we sort the functional data set such that the ordering corresponds to the differential rank of the lincRNA knockdown set. (3) We compute a score S_i as the running average of values from the first position to position i. We then define the enrichment score E as the maximum of the absolute value of S_i for all values of $i > 10$. We require $i > 10$ to avoid small fluctuations in the beginning of the ranked list causing fluctuations in the enrichment score. This score is computed for each lincRNA knockdown by functional set. Because we have many lincRNA knockdowns and functional sets, in reality we have a matrix of scores and we will refer to the enrichment score of the *i*th knockdown and *j*th functional set as E_{ij}.

To assess the significance of these scores, we compute a permutation-derived FDR and assign a confidence value for each projection. Specifically, to assess the significance of E_{ij}, we permute the lincRNA knockdown samples and control samples and compute the enrichment score for each pair across all permutations. To account for the FDR associated with many lincRNAs and functional sets, we use the values of all permutations directly to assess the FDR level of E_{ij}. Specifically, to assess the FDR for each enrichment value E_{ij}, we accumulate all the permutation values for all lincRNA knockdowns and functional sets and compute the number of values greater than E_{ij} as well as a vector of values greater than E_{ij} corresponding to each permutation. The FDR is computed as the average number of permuted values greater than E_{ij} divided by the observed number greater than E_{ij}. Using this approach, we assign an FDR value to each lincRNA knockdown by functional set and identify significant hits as those with an FDR <0.01.

To highlight the accuracy of this approach, we observed that for publicly available gene perturbations for which we also perturbed the gene we were able to identify a significant association of target genes in ~75% of cases. Although the remaining few did not pass our conservative significance criteria, they also showed increased enrichments consistent with their common effects. In addition, the projected effects are highly reproducible across distinct experiments originating from many groups and across multiple expression platforms. Highlighting the specificity of this approach, we note that there are many profiles for which no lincRNA had a similar effect.

Analysis of gene-expression overlaps between independent hairpin knockdowns. To determine whether independent hairpins targeting the same lincRNA gene share common gene targets, we computed a continuous enrichment score described above. Briefly, we computed a *t*-statistic for both hairpins against the negative controls. We then took the second hairpin and sorted the genes. We scored the best hairpin affected genes based on this ranked order. We assessed the significance of this enrichment by permuting the samples and controls and assigned an FDR of the overlap of the expression effect (as described above).

Discrete gene set analysis. Discrete gene sets were analysed using the Gene Set Enrichment Analysis with a slight modification to the scoring procedure to be more analogous to our continuous scoring procedure (described above). Specifically, we computed the average of the expression changes (defined by the *t*-statistic) for all genes within the discrete gene set upon knockdown[63]. Significance was assessed by permuting the control and sample labels and recomputing the average statistic for each permutation. The FDR was assessed off of these values as described above.

Lineage marker gene analysis. We curated lineage marker gene sets from published work and publicly available sources[17,32,63]. We identified lineage marker genes as significantly upregulated using the differential expression criteria outlined above. We validated the expression of these lineage marker genes for a selected set of lineage marker genes using qPCR (as described above) after a 4-day infection. Specifically, we looked at the expression of *Fgf5* (ectoderm), *Sox1* (neuroectoderm), *Sox17* (endoderm), brachyury (mesoderm) and *Cdx2* (trophectoderm). Primer sequences are listed in Supplementary Table 9. Expression estimates were normalized to Gapdh and compared to a panel of 25 negative control hairpins.

Identifying bound lincRNA promoters. We obtained genome-wide transcription factor binding data in mouse ES cells from two sources. The transcription factors Oct4, Sox2, Nanog and Tcf3 were downloaded from the Gene Expression Omnibus (GSE11724) and c-Myc, n-Myc, Zfx, Stat3, Smad1, Klf4 and Esrrb from GEO (GSE11431). For each ChIP-Seq data set, the raw reads were obtained from the SRA (http://www.ncbi.nlm.nih.gov/sra) and processed as follows. (1) The reads were all aligned to the mouse genome assembly (build MM9) using the Bowtie aligner[66], requiring a single best placement of each read. All reads with multiple acceptable placements were removed from the analysis. (2) Binding sites were determined from the aligned reads using the MACS[67] (http://liulab.dfci.harvard.edu/MACS/) algorithm using the default parameters with –mfold 8 to account for varying read counts in the libraries. (3) lincRNA promoter regions were defined as previously described[2,3] using the location of the K4me3 peaks overlapping or within 5 kb of the transcriptional start site determined by RNA-Seq

reconstruction. (4) The transcription factor binding locations and lincRNA promoter locations were intersected and the enrichment level of the peak overlapping a lincRNA promoter was assigned transcription factor binding enrichment for each lincRNA. We defined transcription factor binding locations for protein-coding genes in a comparable way. (5) To exclude the possibility that some of this binding might be due to transcription factor binding at distal enhancers, we excluded all binding events that showed evidence of P300—a protein associated with active enhancers[68]—localization. Altogether, we only identified ~5% of promoters overlapping with any P300 enrichment signal, a slightly lower percentage than identified for protein-coding gene promoters with detectable P300 signal.

Identifying transcription-factor-regulated lincRNA genes. lincRNA probes on the Agilent microarray were analysed using the differential expression methodology described above after knockdown of the transcription factor and comparison to the negative control hairpins. To confirm the expression changes of these lincRNAs, we hybridized 12 transcription factor knockdowns on a custom lincRNA codeset using the Nanostring nCounter assay[41] (LIN-MES1-96). The knockdowns were profiled in biological duplicate along with 15 negative controls. Regulated lincRNAs were identified using the differential expression approach described above.

Nanostring probe-set design. Nanostring probes against lincRNA genes were designed following the standard nanostring design principles with the following modifications specifically for the lincRNA probes. (1) To exclude possible cross-hybridization, probes were screened for cross-hybridization against both the standard mouse transcriptome as well as a background database constructed from all the lincRNA sequences. (2) To account for isoform coverage, a first pass design attempted to select a probe that would target as many isoforms as possible for each lincRNA. In cases where it was not possible to target all isoforms for a given lincRNA, the probe that targeted the largest number was selected, and additional probes were chosen when possible to target the remaining isoforms. (3) The standard restrictions on melting temperature and sequence composition were relaxed to include probes for as many lincRNAs as possible.

Retinoic acid differentiation. V6.5 cells were cultured on gelatin-coated dishes in mES media in the absence of LIF. 5 μM of retinoic acid was added daily and cell samples were taken daily for 6 days. RNA was extracted using Qiagen's RNeasy spin columns following the manufacturer's protocol.

Western blots. 30 μg of mESC nuclear protein extracts were run on 10% Bis-Tris gels (Invitrogen NP0316BOX) in MOPS buffer (Invitrogen NP0001) at 75 V for 20 min followed by 120 V for 1 h. Gels were incubated for 30 min in 20% methanol transfer buffer (Invitrogen NP0006-1) and transferred onto PVDF membranes (Invitrogen 831605) at 20 V for 1 h using the Bio-Rad semi-dry transfer system (170-3940). Membranes were blocked in Blotto (Pierce, 37530) at room temperature for 1 h. Antibodies were diluted in Blotto and membranes were incubated overnight at 4 °C. Antibodies were diluted in the following concentrations. Ezh2 1:2,000, Suz12 1:5,000, hnRNPH 1:1,000, Ruvbl2 1:1,000, Jarid1b 1:500, Hdac1 1:250, Cbx6 1:500, Yy1 1:500. All antibodies tested were raised in rabbit. The next day, membranes were washed 3× in 0.1% TBST for 5 min each. The membranes were probed with anti-rabbit-horse radish peroxidase (GE Healthcare; NA9340V) at a 1:10,000 dilution, washed 3× in 0.1% TBST, incubated in ECL reagent (GE Healthcare RPN2132) and exposed.

Crosslinked RNA immunoprecipitation. V6.5 mES cells were fixed with 1% formaldehyde for 10 min at room temperature, quenched with 2.5 M glycine, washed with 1× PBS (3×) harvested by scraping, pelleting, and re-suspended in modified RIPA lysis buffer (150 mM NaCl, 50 mM Tris, 0.5% sodium deoxycholate, 0.2% SDS, 1% NP-40) supplemented with RNase inhibitors (Ambion, AM2694) and protease inhibitors. For UV crosslinking experiments, cells were irradiated with 254 nm UV light. Cells were kept on ice and crosslinked in 1× PBS using 400,000 μjoules cm^{-2}.

Cell suspension was sonicated using a Branson 250 Sonifier for 3 × 20 s cycles at 20% amplitude. 10 μl of Turbo DNase (Ambion, AM2238) was added to sonicated material, incubated at 37 °C for 10 min, and spun down at max speed for 10 min at 4 °C. Protein-G beads were washed and pre-incubated with antibodies for 30 min at room temperature. Lysate and beads were incubated at 4 °C for 2 h. Beads were washed 3× using the following wash buffer (1× PBS, 0.1% SDS, 0.5% NP-40) followed by 2× using a high salt wash buffer (5× PBS, 0.1% SDS, 0.5% NP-40) and crosslinks were reversed and proteins were digested with 5 μl proteinase-K (NEB, P8102S) at 65 °C for 2–4 h. RNA was purified using phenol/chloroform/isoamyl alcohol and RNA was precipitated in isopropanol.

Nanostring hybridization. 500 ng of total RNA was hybridized for 17 h using the lincRNA code set. The hybridized material was loaded into the nCounter prep station followed by quantification on the nCounter Digital Analyser following the manufacturer's protocol. For RNA immunoprecipitation experiments, we used a modified protocol. After reverse crosslinking, RNA was extracted using phenol/chloroform and ethanol precipitation methods and re-suspended in 10 μl of H$_2$O. 5 μl of the eluted material was hybridized for 17 h using the lincRNA code set.

Nanostring analysis. Probe values were normalized to negative control probes by dividing the value of the probe by the maximum negative control probe. Probe values were floored to a normalized value of 3 (threefold higher than maximum negative control). Probes with no value greater than this floor across all samples were removed from the analysis. The values were log transformed. To control for variability between runs and different input material amounts, we normalized all samples simultaneously using the quantile normalization approach described above. The result is a set of normalized log-expression values for each probe normalized across all experiments.

Validation of RNA immunoprecipitation methods. To validate our formaldehyde-based RNA immunoprecipitation method we immunoprecipitated the RNA binding protein hnRNPH, which has a role in mRNA splicing[69] and identified the associated RNAs. Consistent with known interactions, we identified a strong enrichment for its binding to intronic regions of mRNA genes. We validated these observed results in mouse ES cells by performing UV-crosslinking experiments[70–72] and identified nearly identical results. We identified a similar correlation between the UV and formaldehyde crosslinked samples as for biological replicates of UV crosslinked samples and formaldehyde crosslinked samples and highly comparable enrichments (data not shown).

Antibody selection. We selected chromatin proteins that have been implicated in regulation of the pluripotent state along with their known associated 'reader', 'writer' and 'eraser' complexes. Specifically, we tested antibodies against 40 chromatin proteins, corresponding to 28 chromatin complexes. In many cases, we tested multiple antibodies against the same target protein to try to identify an antibody that worked well for immunoprecipitation. A full list of tested complexes and their associated antibodies is listed in Supplementary Table 18.

Determining significant chromatin–lincRNA enrichments. We tested each antibody using formaldehyde crosslinked cells and had a two-step procedure for considering an antibody successful. (1) We tested all selected antibodies in batches, with each batch containing a mock-IgG (Santa Cruz) negative control and hnRNPH (Bethyl) positive control. Batches with variability in either the mock-IgG or hnRNPH controls were excluded and retested. For each successful batch, we computed enrichment for each lincRNA between the tested antibody and mock-IgG. We considered an antibody successful in the first step if the highest enrichment level exceeded a fivefold change compared to the mock-IgG control and more than 10 lincRNAs exceeded this threshold. Although this approach can yield false positives (antibodies that pass but are not efficient) it significantly reduced the number of antibodies to be tested in the next step. (2) For all antibodies that successfully passed the first criterion, we performed immunoprecipitation on two additional biological replicates along with 4 mock-IgG controls. We computed a t-statistic for each lincRNA compared to the controls and assessed the significance using a permutation test, by permuting the samples and IgG samples (as above). Hits were considered significant if they exceed a t-statistic cutoff of 2 (log scale) compared to the controls and had an FDR <0.2. We allowed a slightly higher FDR cutoff because the number of permutations was far smaller yielding lower power to estimate the FDR. Only antibodies yielding significant lincRNAs were considered successful. In total, we identified 12 of the 28 complexes (55 antibodies) with at least one successful antibody.

Determining significant overlaps between lincRNA and chromatin protein knockdown effects. To determine the functional overlap between the lincRNA and the chromatin complexes it physically interacts with, we compared the effects on gene expression upon knockdown of the lincRNA and the associated protein complex. To do this, we used the gene expression profiles determined for each lincRNA knockdown and knockdowns of 9 of the 12 identified chromatin complexes for which we had good hairpins. We defined each interaction between a lincRNA and protein, and computed a continuous enrichment score, generated all permutations of the control hairpins and sample hairpins and assigned an FDR to the scores (as described above). At an FDR <0.05 we identified 43% of the interactions to be significant. For 69% of the interactions, we were able to identify an overlap at an FDR <0.1.

47. Meissner, A., Eminli, S. & Jaenisch, R. Derivation and manipulation of murine embryonic stem cells. *Methods Mol. Biol.* **482,** 3–19 (2009).
48. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95,** 379–391 (1998).
49. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17,** 126–140 (2003).
50. Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* **12,** 2048–2060 (1998).
51. Nakatake, Y. *et al.* Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. *Mol. Cell. Biol.* **26,** 7772–7782 (2006).
52. Brons, I. G. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448,** 191–195 (2007).

53. Torres-Padilla, M. E., Parfitt, D. E., Kouzarides, T. & Zernicka-Goetz, M. Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature* **445,** 214–218 (2007).
54. Gaspar-Maia, A. *et al.* Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* **460,** 863–868 (2009).
55. Dejosez, M. *et al.* Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. *Cell* **133,** 1162–1174 (2008).
56. Yuan, P. *et al.* Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes Dev.* **23,** 2507–2520 (2009).
57. Pruitt, K. D., Tatusova, T., Klimke, W. & Maglott, D. R. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res.* **37,** D32–D36 (2009).
58. Yang, Y. H. *et al.* Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30,** e15 (2002).
59. Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* **32,** 491–502 (2008).
60. Aiba, K. *et al.* Defining a developmental path to neural fate by global expression profiling of mouse embryonic stem cells and adult neural stem/progenitor cells. *Stem Cells* **24,** 889–895 (2006).
61. Ying, Q. L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nature Biotechnol.* **21,** 183–186 (2003).
62. Morrisey, E. E. *et al.* GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev.* **12,** 3579–3590 (1998).
63. Bock, C. *et al.* Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144,** 439–452 (2011).
64. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462,** 108–112 (2009).
65. Lamb, J. *et al.* The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313,** 1929–1935 (2006).
66. Langmead, B., Hansen, K. D. & Leek, J. T. Cloud-scale RNA-sequencing differential expression analysis with Myrna. *Genome Biol.* **11,** R83 (2010).
67. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
68. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457,** 854–858 (2009).
69. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods* **7,** 1009–1015 (2010).
70. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456,** 464–469 (2008).
71. Ule, J. *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302,** 1212–1215 (2003).
72. Wang, Z., Tollervey, J., Briese, M., Turner, D. & Ule, J. CLIP: construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins *in vivo*. *Methods* **48,** 287–293 (2009).

LETTER

Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus

A list of authors and their affiliations appears at the end of the paper

Both obesity and being underweight have been associated with increased mortality[1,2]. Underweight, defined as a body mass index (BMI) ≤ 18.5 kg per m^2 in adults and ≤ -2 standard deviations from the mean in children, is the main sign of a series of heterogeneous clinical conditions including failure to thrive[3–5], feeding and eating disorder and/or anorexia nervosa[6,7]. In contrast to obesity, few genetic variants underlying these clinical conditions have been reported[8,9]. We previously showed that hemizygosity of a ~600-kilobase (kb) region on the short arm of chromosome 16 causes a highly penetrant form of obesity that is often associated with hyperphagia and intellectual disabilities[10]. Here we show that the corresponding reciprocal duplication is associated with being underweight. We identified 138 duplication carriers (including 132 novel cases and 108 unrelated carriers) from individuals clinically referred for developmental or intellectual disabilities (DD/ID) or psychiatric disorders, or recruited from population-based cohorts. These carriers show significantly reduced postnatal weight and BMI. Half of the boys younger than five years are underweight with a probable diagnosis of failure to thrive, whereas adult duplication carriers have an 8.3-fold increased risk of being clinically underweight. We observe a trend towards increased severity in males, as well as a depletion of male carriers among non-medically ascertained cases. These features are associated with an unusually high frequency of selective and restrictive eating behaviours and a significant reduction in head circumference. Each of the observed phenotypes is the converse of one reported in carriers of deletions at this locus. The phenotypes correlate with changes in transcript levels for genes mapping within the duplication but not in flanking regions. The reciprocal impact of these 16p11.2 copy-number variants indicates that severe obesity and being underweight could have mirror aetiologies, possibly through contrasting effects on energy balance.

Copy-number variants (CNVs) at the 16p11.2 locus have been associated with cognitive disorders including autism (deletions) and schizophrenia (duplications)[11–13], conditions that have been suggested to lie at opposite ends of a single spectrum of psychiatric phenotypes[14]. We and others have reported that a deletion of this region spanning 28 genes (Supplementary Table 1) increases the risk of morbid obesity 43-fold (Supplementary Fig. 1)[10,15]. We hypothesized that the reciprocal duplication, with its resulting increase in gene dosage, may influence BMI in a converse manner. The duplication was identified in 73 out of 31,424 patients with DD/ID, a frequency consistent with previous reports[13] (Table 1). Four additional cases were identified among 1,080 patients affected by bipolar disease or schizophrenia. Compared to its prevalence in seven European population-based genome-wide association study (GWAS) cohorts[16–18] (31 out of 58,635 individuals), the duplication was significantly more frequent in both the DD/ID cohorts ($P = 4.23 \times 10^{-13}$; odds ratio = 4.4, 95% confidence interval = 2.9–6.9) and the psychiatric cohorts ($P = 3.6 \times 10^{-3}$; odds ratio = 7.0, 95% confidence interval = 1.8–19.9) (Table 1), strengthening previous reports of similar associations[12,13]. Our data do not support a two-hit model[19] for the effects of 16p11.2 duplications or deletions (Supplementary Text and Supplementary Table 2).

We compared available data on weight, height and BMI for 106 independent duplication carriers (including published cases) to data for reference populations matched for gender, age and geographical location (Table 2, Methods and Supplementary Tables 3 and 4). The duplication was strongly associated with lower weight (mean Z-score -0.56; $P = 4.4 \times 10^{-4}$) and lower BMI (mean Z-score -0.47; $P = 2.0 \times 10^{-3}$) (Table 2 and Supplementary Table 5). Birth parameters ($n = 48$) were normal, indicating a postnatal effect. Adults carrying the duplication had a relative risk of being clinically underweight (BMI <18.5) of 8.3 (95% confidence interval = 4.4–15.9, $P = 1.53 \times 10^{-10}$) (see Methods). Concordantly, none of the 3,544 patients in our obesity cohorts[10,15] carried the duplication (Table 1).

To investigate these associations further, we carried out separate analyses of carrier patients (DD/ID and psychiatric) and non-medically ascertained carriers (population-based cohorts, plus 11 transmitting parents and three other affected first-degree relatives for whom data were available) (Table 2). Each category had significantly lower weight and BMI, with similar effect sizes. However, the proportion of underweight cases (BMI ≤ -2 s.d.) was higher in the first group than in the second group (17 out of 76 compared to 2 out of 40; $P = 0.017$). Note that the impact of the duplication on underweight status might be underestimated here owing to prescription of antipsychotic treatments that are often associated with weight gain[20] (Supplementary Table 6).

Having demonstrated an association of the duplication with being underweight, we investigated the implications of gender for the resulting phenotypes (Fig. 1, Supplementary Fig. 2 and Supplementary Table 7). In DD/ID patients, the impact of the duplication on being underweight is stronger in males; the effect in females is in the same direction, but is smaller and not statistically significant (Table 2). A similar and significant difference ($P = 0.0168$) was observed in adult carriers (all groups combined): the relative risk of being underweight for males is 23.2 (95% confidence interval = 9.1–59.3, $P = 4.6 \times 10^{-11}$); for females it is only 4.7 (95% confidence interval = 1.9–11.8, $P = 9.9 \times 10^{-4}$). A gender bias was also observed in the ascertainment of DD/ID duplication carriers, in which we have an excess of males (51 males:33 females, $P = 0.044$). By contrast, carriers from the general population showed a strong overrepresentation of females (10 males:21 females, $P = 0.035$) (Supplementary Text). A similar bias was observed among transmitting parents (7 males:23 females, $P = 5.53 \times 10^{-4}$). Thus, there is an overrepresentation of males in the medically ascertained group, and a depletion in the non-medically ascertained one. We suggest that males may be more likely than females to present severe phenotypes, and that this may account for the gender bias because severely affected males may be less likely to be recruited to adult population cohorts or to be reproductively successful.

As previously reported[21], the duplication was also associated with reduced head circumference (mean Z-score -0.89; $P = 7.8 \times 10^{-6}$) (Fig. 1), 26.7% presenting with microcephaly (head circumference ≤ -2 s.d.), whereas carriers of the reciprocal deletion had an increased head circumference (mean Z-score $+0.57$; $P = 1.79 \times 10^{-5}$) (Supplementary Fig. 3 and Supplementary Table 8): an additional instance of a mirror phenotype associated with reciprocal changes in copy number at this locus. Notably, head circumference Z-scores correlate

Table 1 | 16p11.2 rearrangements in cases and controls

Ascertainment	Cohorts	Duplication		Deletion		Total
		n	P†	n	P†	
Neuro-developmental disorders	Unspecified DD/ID* from 28 cytogenetic centres	72		113		30,323
	ADHD‡, deCODE	0		1		591
	Childhood autism‡, deCODE	0		2		159
	Childhood autism spectrum disorder‡, deCODE	1		3		351
	TOTAL	**73**	4.23×10^{-13}	**119**	5.43×10^{-32}	**31,424**
	Rearrangement frequency (95% CI)	**0.23%** (0.18–0.29)		**0.38%** (0.31–0.45)		
Family history	First-degree relatives of probands	30		35		43/62‖
Adult psychiatric symptoms	Schizophrenia, deCODE	0		1		657
	Bipolar disease, Rouen	1		0		156
	Schizophrenia, schizo-affective, Rouen	3		0		267
	TOTAL	**4**	3.57×10^{-3}	**1**	3.78×10^{-1}	**1,080**
	Rearrangement frequency (95% CI)	**0.37%** (0.01–0.73)		**0.09%** (0–0.27)		
Underweight	Eating disorder, Spain	1§		0		441
Obesity	Obesity, Spain	0		2		653
	Adult obesity, France	0		4		705
	Childhood obesity, France & UK	0		7		1,574
	Obesity bariatric surgery, France	0		2		141
	Obesity discordant siblings, Sweden	0		2		159
	Obesity and cognitive delay, France & UK	0		9		312
	TOTAL	**0**	4.21×10^{-1}	**26**	2.52×10^{-19}	**3,544**
	Rearrangement frequency (95% CI)	0		**0.73%** (0.45–1.01)		
Population-based cohorts	NFBC1966 Finnish	4		3		5,319
	CoLaus Swiss	5		0		5,612
	EGCUT Estonian	2		1		2,994
	deCODE Iceland	17		18		36,601
	SHIP Germany	1		2		4,070
	KORA F3+F4 Germany	2		1		3,458
	Paediatric family study	0		0		581
	TOTAL	**31**		**25**		**58,635**
	Rearrangement frequency (95% CI)	**0.05%** (0.03–0.07)		**0.04%** (0.03–0.06)		

CI, Confidence interval; ADHD, attention-deficit hyperactivity disorder. *Not a disease-specific cohort. Detailed distribution is provided in the online methods. †Fisher's exact test, compared to the combined frequency in general population groups. ‡There was no overlap between these 3 cohorts. §Atypical duplication (see Supplementary Fig. 5). ‖ Total number of parental pairs tested for duplication/deletion. 13 out of 43 duplications and 27 out of 62 deletion cases were de novo.

positively with those of BMI in carriers of both the duplication (rho = 0.37; $P = 2.65 \times 10^{-3}$) and the deletion (rho = 0.42; $P = 1.9 \times 10^{-5}$) (Supplementary Methods). This indicates that head circumference and BMI may be regulated by a common pathway, or that a causal relationship exists between these two traits in these patients. Alternatively, the two phenotypes may arise from distinct genes and pathways. A full list of malformations and secondary phenotypes reported in duplication carriers ascertained for DD/ID is available in Supplementary Table 9.

In view of the importance of modified eating behaviours in obesity and being underweight, the clinical reports of duplication carriers were screened for evidence of such modified behaviours. In 11 out of 77 clinically ascertained cases, clinicians had spontaneously reported low food intake and selective and restrictive eating behaviour, again mirroring one of the phenotypes—hyperphagia—seen in deletion carriers[10] (Supplementary Table 6) and indicating that the duplication may increase the risk of eating disorders. Consequently, we carried out

multiplex ligation-dependent probe amplification (MLPA, Supplementary Table 10) to screen for 16p11.2 rearrangements in 441 patients diagnosed with eating disorders, including anorexia nervosa, bulimia and binge eating disorder (Table 1 and Supplementary Text). No duplications of the entire region were identified, but one out of 109 anorexia nervosa patients carried an atypical 136-kb duplication that encompasses the sialophorin (SPN) and quinolinate phosphoribosyltransferase (QPRT) genes (Supplementary Fig. 4). This single, smaller duplication does not allow us to draw any firm conclusions, but together with other atypical rearrangements, it may, in the future, be essential for establishing the roles of the 28 genes within the region.

Large genomic structural variants are known to affect the expression of genes not only within the affected region but also at a distance[22–25]. Therefore, it is possible that the phenotypes observed in 16p11.2 deletion and duplication individuals are due to effects on the expression of genes mapping outside the rearranged interval, rather than to gene dosage within the 600-kb deletion or duplication. We measured

Table 2 | Comparisons of the height, weight and BMI distributions in duplication carriers and controls.

	Strata	Combined†			DD/ID or psychiatric†			Non-medically ascertained‡		
		Mean Z-score	P-value	n*	Mean Z-score	P-value	n*	Mean Z-score	P-value	n*
BMI	All	−0.47	**2.0×10^{-3}**	102	−0.56	**4.1×10^{-3}**	76	−0.45	**6.0×10^{-3}**	40
	Male	−0.54	**2.1×10^{-2}**	52	−0.71	**1.3×10^{-2}**	43	−0.31	2.0×10^{-1}	14
	Female	−0.4	**1.8×10^{-2}**	50	−0.37	8.3×10^{-2}	33	−0.52	**4.2×10^{-3}**	26
Weight	All	−0.56	**4.4×10^{-4}**	104	−0.65	**1.3×10^{-3}**	78	−0.61	**3.0×10^{-3}**	40
	Male	−0.64	**5.8×10^{-3}**	53	−0.79	**4.4×10^{-3}**	44	−0.57	8.8×10^{-2}	14
	Female	−0.47	**1.7×10^{-2}**	51	−0.47	6.5×10^{-2}	34	−0.63	**8.6×10^{-3}**	26
Height	All	−0.24	4.8×10^{-2}	103	−0.33	3.6×10^{-2}	77	−0.15	1.8×10^{-1}	40
	Male	−0.34	4.5×10^{-2}	52	−0.4	4.6×10^{-2}	43	−0.29	1.2×10^{-1}	14
	Female	−0.14	2.6×10^{-1}	51	−0.24	2.1×10^{-1}	34	−0.07	3.7×10^{-1}	26

The available BMI, weight and height data for duplication carriers were transformed to Z-scores using gender- and age-matched reference populations, and one-tailed t-tests were carried out to determine whether the mean Z-scores deviated from zero. Significant differences were identified by reference to cutoffs controlling the false discovery rate at 5% (see Methods): BMI, 0.022; weight, 0.032; height, 0.025. Significant results are indicated in bold. Data were not available for all subjects. *Relatives of probands were excluded as required, to avoid including more than one member of the same family in a single analysis. †Including 24 cases from the literature (Supplementary Table 3). ‡Population-based cases and first-degree relatives of probands.
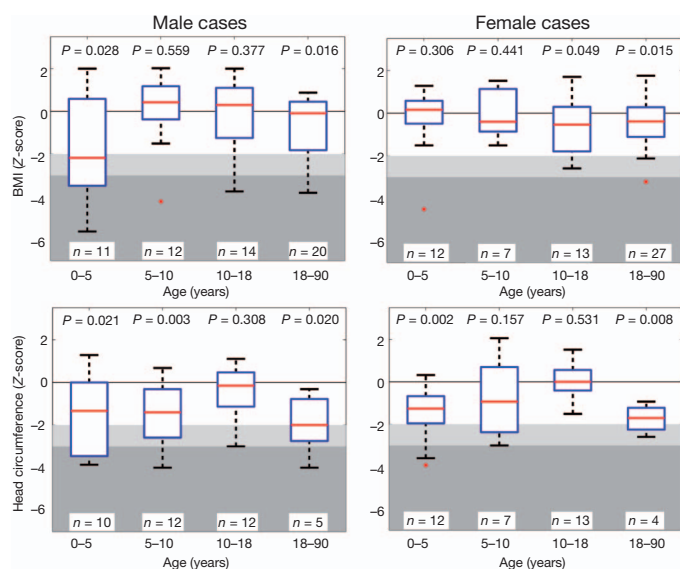
Figure 1 | Effect of the chromosome 16p11.2 duplication on BMI and head circumference. *Z*-score values of BMI and head circumference in carriers of the 16p11.2 duplication, stratified by gender and age group. The most severe effect is observed in children at 0–5 years of age. Boxplots represent the fifth, twenty-fifth, median, seventy-fifth and ninety-fifth percentile for each age group. Light grey and dark grey backgrounds represent ≤ −2 and ≤ −3 s.d., respectively, corresponding to the WHO definition of moderately and severely underweight[30]. BMI is decreased in adolescent and adult females.

relative transcript levels of 27 genes mapping within or near to the rearrangement, using lymphoblastoid cell lines (Supplementary Tables 1 and 11): six from deletion carriers, five from duplication carriers and ten from gender- and age-matched controls (Supplementary Table 12). Expression levels correlated positively with gene dosage for all genes in the CNV region (Fig. 2), consistent with published partial results from adipose tissue[10]. Mean relative transcript levels in deletion and duplication carriers were, respectively, 67% and 214% of the levels measured in controls (Supplementary Table 13). Although genes proximal (centromeric) to the rearrangement interval

showed no significant variation in relative transcript levels between patients and controls (Fig. 2), distal (telomeric) genes showed a marked alteration in relative expression. However, their expression levels, including that of *SH2B1* (for which gene dosage and a nearby single nucleotide polymorphism (SNP) have been associated with obesity[15,26]), were similarly upregulated in cell lines of both deletion and duplication carriers, showing no apparent correlation between transcript level and either copy number or phenotype (Fig. 2). Although lymphoblastoid cells may not recapitulate obesity-relevant tissues, previous experiments have shown a high degree of correlation between expression levels in different tissues and cell lines[22], indicating that the same pathways may be similarly disrupted in different cell lineages. Thus, any involvement of these distal genes in the control of BMI in these subjects seems unlikely.

Our study demonstrates the power of very large screens (>95,000 samples: to our knowledge the largest of its kind so far) to characterize the clinical and molecular correlates of a rare functional genomic variant. We demonstrate unambiguously that carrying the 16p11.2 duplication confers a high risk of being clinically underweight, and show that reciprocal changes in gene dosage at this locus result in several mirror phenotypes. As in the schizophrenia/autism[14] and microcephaly/macrocephaly[21] dualisms, abnormal eating behaviours, such as hyperphagia and anorexia, could represent opposite pathological manifestations of a common energy-balance mechanism, although the precise relationships between these mirror phenotypes remain to be determined. We speculate that head circumference (which correlates with brain volume[27]), and thus neuronal circuitry, may affect cognitive function and energy balance in patients with 16p11.2 rearrangements (possibly through eating behaviour). Consistent with this are previous reports that a subgroup of children with microcephaly show a concomitant reduction in weight percentile[28]. Our findings also support the observation that severe overweight and underweight phenotypes correlate with lower cognitive functioning[4,29]. Thus, abnormal food intake may be a direct result of particular neurodevelopmental disorders. Although it is possible that the 16p11.2 region encodes distinct genes specific for each trait, a more parsimonious hypothesis is that these clinical manifestations of dysfunction of the central nervous system are all secondary to the disruption of a single neurodevelopmental step that is sensitive to gene dosage. Further resolution of this issue may
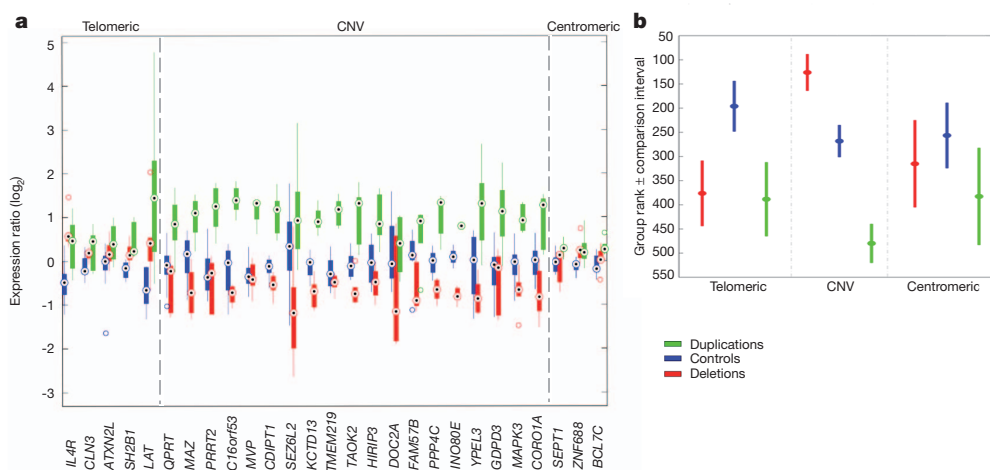
Figure 2 | Transcript levels for genes within and near to the 16p11.2 rearrangements. a, Relative expression levels of 27 genes mapping to 16p11.2 in deletion (*n* = 6) and duplication (*n* = 5) carriers (red and green, respectively), and in control cell lines (*n* = 10, blue). Grey lines denote the extent of the 16p11.2 CNV (29.5–30.1 megabases (Mb)). Complete lists of genes mapping within the rearranged interval, and of the quantitative PCR assays, are in Supplementary Tables 1 and 11, respectively. For the possible relevance of each of these genes to obesity/leanness and/or developmental delay/cognitive deficits, see ref. 10. **b**, Rank comparison (Kruskal–Wallis test)

between the expression of 27 genes mapping to 16p11.2 in deletion and duplication carriers (red and green, respectively) and in control cell lines (blue). Genes are labelled as telomeric, centromeric or within the rearranged interval (CNV). Dots correspond to the mean group rank and bars indicate the comparison interval. Groups with non-overlapping intervals are significantly different (*P*-values were adjusted for multiple testing issues using a Bonferroni correction, where the number of tests is the number of pairwise comparisons; the resulting adjusted *P*-value was less than 0.05).

require the identification of additional patients with rare atypical rearrangements in this region.

METHODS SUMMARY

Underweight is defined in adults as BMI ≤ 18.5. In individuals younger than 18 years of age, it is defined as a Z-score ≤ −2.

Statistics. Two-tailed Fisher's exact test was used to compare frequencies of the rearrangement in patients and controls. Z-scores were computed for all data using gender-, age- and geographically-matched reference populations. One-tailed Student's t-test was performed to test BMI, height, weight and head circumference in duplication carriers for Z-scores of less than zero. We used Kruskal–Wallis tests for differences in gene expression patterns. P-values were adjusted using a Bonferroni correction, considering the number of pairwise comparisons; the resulting adjusted P-value was less than 0.05. The relative risk of being underweight was calculated as the ratio of the fraction of underweight individuals among duplication carriers versus our control group.

Discovery of CNVs. Carriers of 16p11.2 duplication and deletion were identified through various procedures: (1) comparative genomic hybridization with Agilent 44K, 60K, 105K, 180K, 244K arrays; (2) Illumina Human317, Human370, HumanHap550, Human610 and 1M BeadChips; (3) Affymetrix 6.0, 500K genotyping arrays; (4) quantitative multiplex PCR of short fluorescent fragments (QMPSF); (5) fluorescent in situ hybridization (FISH); (6) MLPA. CNV analyses of GWAS data were carried out using cnvHap, a moving-window average-intensity procedure, a Gaussian mixture model, circular binary segmentation, QuantiSNP, PennCNV, BeadStudio GT module and Birdseed. At least two independent algorithms were used for each cohort.

Expression analyses. Lymphoblastoid cell lines were established from carriers and controls. SYBR Green quantitative PCR was performed to assess relative expression of genes.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Berrington de Gonzalez, A. et al. Body-mass index and mortality among 1.46 million white adults. N. Engl. J. Med. **363**, 2211–2219 (2010).
2. Flegal, K. M., Graubard, B. I., Williamson, D. F. & Gail, M. H. Cause-specific excess deaths associated with underweight, overweight, and obesity. J. Am. Med. Assoc. **298**, 2028–2037 (2007).
3. Olsen, E. M. et al. Failure to thrive: the prevalence and concurrence of anthropometric criteria in a general infant population. Arch. Dis. Child. **92**, 109–114 (2007).
4. Corbett, S. S. & Drewett, R. F. To what extent is failure to thrive in infancy associated with poorer cognitive development? A review and meta-analysis. J. Child Psychol. Psychiatry **45**, 641–654 (2004).
5. Rudolf, M. C. & Logan, S. What is the long term outcome for children who fail to thrive? A systematic review. Arch. Dis. Child. **90**, 925–931 (2005).
6. Bravender, T. et al. Classification of eating disturbance in children and adolescents: proposed changes for the DSM-V. Eur. Eat. Disord. Rev. **18**, 79–89 (2010).
7. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR 39–134 and 583–596 (Am. Psychiatric Assoc., 2000).
8. Scherag, S., Hebebrand, J. & Hinney, A. Eating disorders: the current status of molecular genetic research. Eur. Child Adolesc. Psychiatry **19**, 211–226 (2010).
9. Wang, K. et al. A genome-wide association study on common SNPs and rare CNVs in anorexia nervosa. Mol. Psychiatry doi:10.1038/mp.2010.107 (2010).
10. Walters, R. G. et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. Nature **463**, 671–675 (2010).
11. Marshall, C. R. et al. Structural variation of chromosomes in autism spectrum disorder. Am. J. Hum. Genet. **82**, 477–488 (2008).
12. McCarthy, S. E. et al. Microduplications of 16p11.2 are associated with schizophrenia. Nature Genet. **41**, 1223–1227 (2009).
13. Weiss, L. A. et al. Association between microdeletion and microduplication at 16p11.2 and autism. N. Engl. J. Med. **358**, 667–675 (2008).
14. Crespi, B., Stead, P. & Elliot, M. Evolution in health and medicine Sackler colloquium: comparative genomics of autism and schizophrenia. Proc. Natl Acad. Sci. USA **107** (Suppl 1), 1736–1741 (2010).
15. Bochukova, E. G. et al. Large, rare chromosomal deletions associated with severe early-onset obesity. Nature **463**, 666–670 (2010).
16. Firmann, M. et al. The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. BMC Cardiovasc. Disord. **8**, 6 (2008).
17. Sabatti, C. et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nature Genet. **41**, 35–46 (2009).
18. Nelis, M. et al. Genetic structure of Europeans: a view from the north-east. PLoS ONE **4**, e5472 (2009).
19. Girirajan, S. & Eichler, E. E. Phenotypic variability and genetic susceptibility to genomic disorders. Hum. Mol. Genet. **19**, R176–R187 (2010).
20. Pramyothin, P. & Khaodhiar, L. Metabolic syndrome with the atypical antipsychotics. Curr. Opin. Endocrinol. Diabetes Obes. **17**, 460–466 (2010).
21. Shinawi, M. et al. Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioral problems, dysmorphism, epilepsy, and abnormal head size. J. Med. Genet. **5**, 332–341 (2009).
22. Merla, G. et al. Submicroscopic deletion in patients with Williams–Beuren syndrome influences expression levels of the nonhemizygous flanking genes. Am. J. Hum. Genet. **79**, 332–341 (2006).
23. Stranger, B. E. et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science **315**, 848–853 (2007).
24. Henrichsen, C. N. et al. Segmental copy number variation shapes tissue transcriptomes. Nature Genet. **41**, 424–429 (2009).
25. Ricard, G. et al. Phenotypic consequences of copy number variation: insights from Smith–Magenis and Potocki–Lupski syndrome mouse models. PLoS Biol. **8**, e1000543 (2010).
26. Willer, C. J. et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nature Genet. **41**, 25–34 (2009).
27. Courchesne, E., Carper, R. & Akshoomoff, N. Evidence of brain overgrowth in the first year of life in autism. J. Am. Med. Assoc. **290**, 337–344 (2003).
28. Baxter, P. S., Rigby, A. S., Rotsaert, M. H. & Wright, I. Acquired microcephaly: causes, patterns, motor and IQ effects, and associated growth changes. Pediatrics **124**, 590–595 (2009).
29. Li, Y., Dai, Q., Jackson, J. C. & Zhang, J. Overweight is associated with decreased cognitive functioning among school-age children and adolescents. Obesity (Silver Spring) **16**, 1809–1815 (2008).
30. de Onis, M., Blossner, M., Borghi, E., Frongillo, E. A. & Morris, R. Estimates of global prevalence of childhood underweight in 1990 and 2015. J. Am. Med. Assoc. **291**, 2600–2606 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Sébastien Jacquemont[1]*, Alexandre Reymond[2]*, Flore Zufferey[1], Louise Harewood[2], Robin G. Walters[3], Zoltán Kutalik[4,5], Danielle Martinet[1], Yiping Shen[6,7], Armand Valsesia[4,5,8], Noam D. Beckmann[1], Gudmar Thorleifsson[9], Marco Belfiore[1], Sonia Bouquillon[10], Dominique Campion[11], Nicole de Leeuw[12], Bert B. A. de Vries[12], Tõnu Esko[13,14], Bridget A. Fernandez[15], Fernando Fernández-Aranda[16], José Manuel Fernández-Real[17], Mònica Gratacòs[18], Audrey Guilmatre[11], Juliane Hoyer[19], Marjo-Riitta Jarvelin[20,21,22], R. Frank Kooy[23], Ants Kurg[13], Cédric Le Caignec[24], Katrin Männik[13], Orah S. Platt[6], Damien Sanlaville[25], Mieke M. Van Haelst[3,26], Sergi Villatoro Gomez[18], Faida Walha[6], Bai-lin Wu[6,27], Yongguo Yu[6,28], Azzedine Aboura[29], Marie-Claude Addor[1], Yves Alembik[30], Stylianos E. Antonarakis[31], Benoît Arveiler[32,33], Magalie Barth[34], Nathalie Bednarek[35], Frédérique Béna[31], Sven Bergmann[4,5], Mylène Beri[36], Laura Bernardini[37], Bettina Blaumeiser[23], Dominique Bonneau[34], Armand Bottani[31], Odile Boute[38], Han G. Brunner[12], Dorothée Cailley[33], Patrick Callier[39], Jean Chiesa[40], Jacqueline Chrast[2], Lachlan Coin[3], Charles Coutton[41], Jean-Marie Cuisset[42], Jean-Christophe Cuvellier[42], Albert David[24], Bénédicte de Freminville[43], Bruno Delobel[44], Marie-Ange Delrue[32,33], Bénédicte Demeer[45], Dominique Descamps[46], Gérard Didelot[41], Klaus Dieterich[47], Vittoria Disciglio[48,110], Martine Doco-Fenzy[49], Séverine Drunat[29], Bénédicte Duban-Bedu[44], Christèle Dubourg[50,51], Julia S. El-Sayed Moustafa[3], Paul Elliott[52,53], Brigitte H. W. Faas[12], Laurence Faivre[54], Anne Faudet[55], Florence Fellmann[1], Alessandra Ferrarini[2], Richard Fisher[56], Elisabeth Flori[30], Lukas Forer[57], Dominique Gaillard[49], Marion Gerard[29], Christian Gieger[58], Stefania Gimelli[31], Giorgio Gimelli[59], Hans J. Grabe[60], Agnès Guichet[34], Olivier Guillin[11], Anna-Liisa Hartikainen[61], Délphine Heron[62,63], Loyse Hippolyte[1], Muriel Holder[38], Georg Homuth[64], Bertrand Isidor[24], Sylvie Jaillard[50,65], Zdenek Jaros[66], Susana Jiménez-Murcia[16], Géraldine Joly Helas[67], Philippe Jonveaux[36], Satu Kaksonen[68], Boris Keren[55], Anita Kloss-Brandstätter[57], Nine V. A. M. Knoers[26,69], David A. Koolen[12], Peter M. Kroisel[70], Florian Kronenberg[57], Audrey Labalme[25], Emilie Landais[49], Elisabetta Lapi[71], Valérie Layet[72], Solenn Legallic[11], Bruno Leheup[73], Barbara Leube[74], Suzanne Lewis[75], Josette Lucas[65], Kay D. MacDermot[76], Pall Magnusson[77], Christian Marshall[78], Michèle Mathieu-Dramard[45], Mark I. McCarthy[79, 80,81], Thomas Meitinger[82,111], Maria Antonietta Mencarelli[48], Giuseppe Merla[83], Alexandre Moerman[38], Vincent Mooser[84], Fanny Morice-Picard[32,33], Mafalda Mucciolo[48], Matthias Nauck[85], Ndeye Coumba Ndiaye[86], Ann Nordgren[87], Laurent Pasquier[88], Florence Petit[38], Rolph Pfundt[12], Ghislaine Plessis[89], Evica Rajcan-Separovic[90], Gian Paolo Ramelli[91], Anita Rauch[92], Roberto Ravazzolo[93], Andre Reis[19], Alessandra Renieri[48], Cristobal Richart[94], Janina S. Ried[58], Claudine Rieubland[95], Wendy Roberts[96], Katharina M. Roetzer[70], Caroline Rooryck[32,33], Massimiliano Rossi[25], Evald Saemundsen[97], Véronique Satre[41], Claudia Schurmann[64], Engilbert Sigurdsson[98], Dimitri J. Stavropoulos[99], Hreinn Stefansson[9], Carola Tengström[100], Unnur Thorsteinsdóttir[9,101], Francisco J. Tinahones[102], Renaud Touraine[43], Louis Vallée[42], Ellen van Binsbergen[26], Nathalie Van der Aa[23], Catherine Vincent-Delorme[103], Sophie Visvikis-Siest[86], Peter Vollenweider[104], Henry Völzke[105], Anneke T. Vulto-van Silfhout[12], Gérard Waeber[104], Carina Wallgren-Pettersson[106], Robert M. Witwicki[2], Simon Zwolinksi[56], Joris Andrieux[10], Xavier Estivill[18], James F. Gusella[7], Omar Gustafsson[9,107], Andres Metspalu[13,14], Stephen W. Scherer[108], Kari Stefansson[9], Alexandra I. F. Blakemore[3], Jacques S. Beckmann[1,4] & Philippe Froguel[3,109]

[1]Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland. [2]Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland. [3]Department of Genomics of Common Disease, Imperial College London, London W12 0NN, UK. [4]Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland. [5]Swiss Institute of Bioinformatics, University of Lausanne, 1015 Lausanne, Switzerland. [6]Laboratory Medicine, Children's Hospital Boston, Boston, Massachusetts 02115, USA. [7]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [8]Ludwig Institute for Cancer Research, University of Lausanne, 1015 Lausanne, Switzerland. [9]deCODE Genetics, Sturlugata 8, IS-101 Reykjavík, Iceland. [10]Laboratoire de Génétique Médicale, Hopital Jeanne de Flandre, CHRU de Lille, 59000 Lille, France. [11]INSERM U614, University of Rouen and Centre Hospitalier du Rouvray, 76000 Rouen, France. [12]Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands. [13]Institute of Molecular and Cell Biology, University of Tartu, 51010 Tartu, Estonia. [14]Estonian Genome Center, University of Tartu, 51010 Tartu, Estonia. [15]Discipline of Genetics and Medicine, Memorial University of Newfoundland, St John's A1B 3V6, Newfoundland, Canada. [16]Department of Psychiatry, University Hospital of Bellvitge-IDIBELL, Ciber Fisiopatologia Obesidad y Nutrición (CIBEROBN), 08907 Barcelona, Spain. [17]Department of Diabetes, Endocrinology, and Nutrition, Hospital Universitari de Girona Dr. Josep Trueta, Institut d'Investigació Biomèdica de Girona, Ciber Fisiopatologia Obesidad y Nutrición (CIBEROBN), Instituto Salud Carlos III, 17007 Girona, Spain. [18]Genes and Disease Program, Center for Genomic Regulation (CRG-UPF), CIBER en Epidemiología y Salud Pública (CIBERESP), 08003 Barcelona, Catalonia, Spain. [19]Institute of Human Genetics, Friedrich-Alexander University Erlangen-Nuremberg, 91054 Erlangen, Germany. [20]Department of Epidemiology and Biostatistics, Imperial College London, School of Public Health, London W2 1PG, UK. [21]Department of Child and Adolescent Health, National Institute for Health and Welfare, Box 310, 90101 Oulu, Finland. [22]Institute of Health Sciences, University of Oulu, and Biocenter Oulu, University of Oulu, Box 5000, 90014 University of Oulu, Finland. [23]Department of Medical Genetics, University and University Hospital Antwerp, B-2650 Antwerp, Belgium. [24]Service de Génétique Médicale, CHU Nantes, 44093 Nantes, France. [25]Service de Cytogénétique Constitutionnelle, Hospices Civils de Lyon, CHU de Lyon and Neuroscience Research Center, TIGER team, UCBL1, Lyon, F-69000, France. [26]Department of Medical Genetics, University Medical Center, 3584 EA Utrecht, The Netherlands. [27]Children's Hospital and Institutes of Biomedical Science, Fudan University, 200032 Shanghai, China. [28]Shanghai Children's Medical Center, 200127 Shanghai, China. [29]Department of Genetics, APHP-Robert DEBRE University Hospital, 75019 Paris, France. [30]Service de Génétique Médicale, CHU Strasbourg, Hôpital de Hautepierre, 67000 Strasbourg, France. [31]Service of Genetic Medicine, University Hospitals of Geneva, 1205 Geneva, Switzerland. [32]Laboratoire Maladies Rares-Génétique et Métabolisme, Université Bordeaux 2, 33076 Bordeaux, France. [33]Service de Génétique Médicale, CHU de Bordeaux, 33076 Bordeaux, France. [34]Service de Génétique, CHU Angers, 49933 Angers, France. [35]Service Pédiatrie, CHU Hôpital Alix de Champagne, 51100 Reims, France. [36]Laboratoire de génétique, CHU Nancy, 54511 Vandoeuvre les Nancy, France. [37]Mendel Laboratory, IRCCS Casa Solievo della Sofferanza Hospital, 71013 San Giovanni Rotondo, Italy. [38]Service de Génétique Clinique, Hôpital Jeanne de Flandre, CHRU de Lille, 59037 Lille, France. [39]Laboratoire de Cytogénétique, CHU le Bocage, 21070 Dijon, France. [40]Laboratoire de Cytogénétique, CHU Caremeau, 30029 Nimes, France. [41]Laboratoire de Génétique Chromosomique, CHU de Grenoble, BP 217, 38043 Grenoble, France. [42]Service de Neurologie pédiatrique, Hôpital Roger Salengro, CHRU de Lille, 59037 Lille, France. [43]CHU-Hôpital Nord, Service de Génétique, CHU Saint Etienne, 42055 St Etienne, France. [44]Centre de Génétique Chromosomique, Hôpital Saint-Vincent-de-Paul, GHICL, 59160 Lille, France. [45]Service de Génétique Médicale, CHRU Amiens, 80000 Amiens, France. [46]Centre hospitalier de Béthune, 62408 Bethune, France. [47]Service de Génétique Clinique, CHU Grenoble, 38043 Grenoble, France. [48]Medical Genetics, Department of Biotechnology, University of Siena, 53100 Siena, Italy. [49]Service de Génétique, HMB, CHU REIMS, IFR 53, 51092 Reims, France. [50]UMR 6061 CNRS, IFR 140 GFAS, Université de Rennes 1, 35043 Rennes, France. [51]Service de Génétique Moléculaire, CHU, 35033 Rennes, France. [52]Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK. [53]MRC-HPA Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK. [54]Centre de Génétique, CHU Dijon, 21079 Dijon, France. [55]Département de Génétique et Cytogénétique, APHP-GH Pitié-Salpêtrière, 75013 Paris, France. [56]Institute of Human Genetics, International Centre for Life, Newcastle Upon Tyne NE1 4EP, UK. [57]Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, 6020 Innsbruck, Austria. [58]Institute of Genetic Epidemiology, Helmholtz Center Munich, German Research Center for Environmental Health, 85764 Neuherberg, Germany. [59]Laboratorio di Citogenetica, G. Gaslini Institute, 16147 Genova, Italy. [60]Department of Psychiatry and Psychotherapy, Ernst-Moritz-Arndt University Greifswald, 17475 Greifswald and D-18437 Stralsund, Germany. [61]Institute of Clinical Medicine, Department of Obstetrics and Gynecology, University of Oulu, 90570 Oulu, Finland. [62]Embryologie AP-HP, Université Pierre et Marie Curie, 75005 Paris, France. [63]Département de Génétique Cytogénétique, Hôpital Pitié-Salpêtriére, 75013 Paris, France. [64]Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt University Greifswald, D-17487 Greifswald, Germany. [65]Laboratoire de Cytogénétique et Biologie Cellulaire, CHU Rennes, 35033 Rennes, France. [66]Abteilung für Kinder und Jugendheilkunde, Landesklinikum Waldviertel Zwettl, 3910 Zwettl, Austria. [67]Department of Genetics, CHU Rouen, 76183 Rouen, France. [68]The Habilitation Unit of Folkhalsan, Folkhalsan, SF 00250 Helsinki, Finland. [69]Department of Human Genetics, University Medical Center, 3584 EA Utrecht, The Netherlands. [70]Institute of Human Genetics, Medical University of Graz, 8010 Graz, Austria. [71]Medical Genetics Unit, Children's Hospital Anna Meyer, 50139 Firenze, Italy. [72]Unité de Génétique, Groupe Hospitalier du Havre, 76600 le Havre, France. [73]Service de Médecine Infantile III et Génétique Clinique, CHU-Nancy et PRES de l'Université de Lorraine UHP Nancy, 54511 Vandoeuvre les Nancy, France. [74]Institute of Human Genetics and Anthropology, Heinrich-Heine University Hospital Duesseldorf, D-40001 Duesseldorf, Germany. [75]Department of Medical Genetics, The University of British Columbia and Child and Family Research Institute, Vancouver V6H 3N1, Canada. [76]North West Thames Regional Genetics Service, Northwick Park & St Marks Hospital, Harrow HA1 3UJ, UK. [77]Child and Adolescent Psychiatry, Landspitali University Hospital, IS-105 Reykjavík, Iceland. [78]The Centre for Applied Genomics, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [79]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Old Road, Headington, Oxford OX3 7LJ, UK. [80]Wellcome Trust Centre for Human Genetics, University of Oxford, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. [81]Oxford NIHR Biomedical Research Centre, Churchill Hospital, Old Road, Headington, Oxford OX3 7LJ, UK. [82]Institute of Human Genetics,Hemholtz Center Munich, German Research Center for Environmental Health and Institute of Human Genetics, Technical University Munich, 85764 Neuherberg, Germany. [83]Medical Genetics Unit, IRCCS Casa Sollievo della Sofferenza Hospital, 71013 San Giovanni

Rotondo, Italy. [84]Genetics, GlaxoSmithKline R&D, 720 Swedeland Road, King of Prussia, Pennsylvania 19406, USA. [85]Institute of Clinical Chemistry and Laboratory Medicine, Ernst-Moritz-Arndt University Greifswald, D-17475 Greifswald, Germany. [86]Cardiovascular Genetics Research Unit, EA4373, Université Henri Poincaré, 54000 Nancy, France. [87]Department of Molecular Medicine and Surgery, Karolinska Institutet, 171 76 Stockholm, Sweden. [88]Service de Génétique-CLAD Ouest, 35203 Rennes cedex 2, France. [89]Service de Génétique, CHU Clémenceau, 14033 Caen, France. [90]Department of Pathology, University of British Columbia and Child and Family Research Institute, Vancouver, British Columbia V5Z 4H4, Canada. [91]Division of Pediatrics, Ospedale San Giovanni, 6500 Bellinzona, Switzerland. [92]Institute of Medical Genetics, University of Zurich, 8603 Schwerzenbach, Switzerland. [93]Department of Pediatrics and CEBR, G. Gaslini Institute, University of Genova, 16126 Genova, Italy. [94]Department of Internal Medicine, University Hospital Juan XXIII, Universitat Rovira y Virgili, Ciber Fisiopatologia Obesidad y Nutrición (CIBEROBN), Instituto Salud Carlos III, 43005 Tarragona, Spain. [95]Division of Human Genetics, Department of Paediatrics, Inselspital, University of Bern, 3010 Bern, Switzerland. [96]The Autism Research Unit, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [97]State Diagnostic and Counseling Center, 200 Kopavogur, Iceland. [98]University of Iceland and Landspitali University Hospital, 101 Reykjavik,

Iceland. [99]Department of Pediatric Laboratory Medicine, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [100]Genetic Services, Rinnekoti Research Foundation, Kumputie 1, SF-02980 Espoo, Finland. [101]Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland. [102]Department of Endocrinology and Nutrition, Clinic Hospital of Virgen de la Victoria, Ciber Fisiopatologia Obesidad y Nutrición (CIBEROBN), Instituto Salud Carlos III, 29010 Malaga, Spain. [103]Centre de Maladies Rares, Anomalies du Developpement Nord de France, CH Arras-CHRU Lille, 59000 Lille, France. [104]Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland. [105]Institute for Community Medicine, Ernst-Moritz-Arndt University Greifswald, D-17475 Greifswald, Germany. [106]Department of Medical Genetics, Haartman Institute, University of Helsinki and Folkhälsan Institute of Genetics, 00251 Helsinki, Finland. [107]Department of Psychiatry, Oslo University Hospital, N-0407 Oslo, Norway. [108]The Hospital for Sick Children, University of Toronto, Toronto, Ontario M5G IL7, Canada. [109]CNRS 8090-Institute of Biology, Pasteur Institute, 59800 Lille, France. [110]UOC Genetica Medica, Azienda Ospedaliera Universitaria Senese, Siena, Italy. [111]Institute of Human Genetics, Technische Universität München, Klinikum rechts der Isar, 81675 München, Germany.

*These authors contributed equally to this work.

METHODS

Study cohorts. For the description of these cohorts, refer to Supplementary Information.

CNV detection. Cases ascertained for intellectual disabilities and developmental delay were identified through standard medical diagnostic procedures. CNV analyses of GWAS data were variously carried out using cnvHap[31]; a moving-window average-intensity procedure; a Gaussian mixture model (Valsesia *et al.*, submitted); circular binary segmentation[32,33]; QuantiSNP[34]; PennCNV[35]; BeadStudio GT module (Illumina Inc.); and Birdseed[36] (see below). At least two independent algorithms were used for each cohort.

Patients referred for intellectual disabilities and developmental delay. All diagnostic procedures (CGH, quantitative PCR and/or quantitative multiplex PCR of short fluorescent fragments) were carried out according to the relevant guidelines of good clinical laboratory practice for the respective countries. All rearrangements in probands were confirmed by a second independent method and karyotyping was performed in all cases to exclude a complex rearrangement.

Northern Finland 1966 birth cohort (NFBC). CNV calling has been previously described[10]. In brief, data were normalized using Illumina BeadStudio, then GC effects on ratios were removed by regressing on GC and GC2, and wave effects were removed by fitting a Loess function[37]. CNV analysis was done using cnvHap[31]. All called 16p11.2 duplications were validated by direct analysis of $\log_2$ ratios. Data for each probe were normalized by first subtracting the median value across all samples (so that the distribution of ratios for each probe was centred on zero), and then dividing by the variance across all samples (to correct for variation in the sensitivity of different probes to copy-number variation). All CNV calls were confirmed by MLPA.

deCODE genetics. Illumina Human317, Human370, HumanHap550, Human610 and 1M BeadChips were used for CNV analysis. BeadStudio (version 2.0) was used to call genotypes, normalize the signal intensity data and establish the log R ratio (LRR) and B allele frequency (BAF) at every SNP according to standard Illumina protocols. All samples passed a standard SNP-based quality control procedure with a SNP call rate greater than 0.97. PennCNV[35], a free, open-source tool, was used for detection of CNVs. The input data for PennCNV are LRR, a normalized measure of the total signal intensity for the two alleles of the SNP, and BAF, a normalized measure of the allelic intensity ratio of the two alleles. These values are derived with the help of control genotype clusters (HapMap samples), using the Illumina BeadStudio software. PennCNV employs a hidden Markov model to analyse the LRR and BAF values across the genome. CNV calls are made on the basis of the probability of a given copy state at the current marker, as well as on the probability of observing a copy-state change from the previous marker to the current one. PennCNV uses a built-in correction model for GC content[38].

Cohorte Lausannoise (CoLaus). Data normalization and CNV calling have been previously described[10]. Data normalization included allelic cross-talk calibration[39,40], intensity summarization using robust median average, and correction for any PCR amplification bias. Wave effects were corrected by fitting a Loess function[37]. CNV calling was done using a Gaussian mixture model (Valsesia *et al.*, submitted) that fits four components (deletion, copy-neutral, one additional copy and two additional copies) to copy-number ratios. The final copy number at each probe location is determined as the expected (dosage) copy number. The method has been validated by comparing test data sets with results from the CNAT[41] and CBS[32,33] algorithms, and by replicating a subset of CoLaus subjects on Illumina arrays. Only duplications found by both Gaussian mixture model and CBS were considered.

Estonian genome center of the University of Tartu (EGCUT). Genotypes were called by BeadStudio software GT module v3.1 or GenomeStudio GT v1.6 (Illumina Inc.). Values for LRR and BAF produced by BeadStudio were formatted for further CNV analysis and break-point mapping with Hidden-Markov-Model-based softwares QuantiSNP (ver.1.1)[34] and PennCNV[42] or CNVPartition 2.4.4 (Illumina Inc.). All analyses were carried out using the recommended settings, except changing EMiters to 25 and L to 1,000,000 in QuantiSNP. For PennCNV, the Estonian-population-specific SNP allele frequency data was used. All detected duplications were confirmed by quantitative PCR.

Study of health in Pomerania (SHIP). Raw intensities were normalized using Affymetrix power tools (Affymetrix); CNV analysis was done using Birdseye from the Birdsuite software package[36] and PennCNV[35]. PennCNV predictions with confidence scores less than 10 were removed. Birdsuite predictions were filtered as in ref. 15: CNVs were kept if their linkage disequilibrium (LOD) score was >10, length >1 kb, number of probes ≥5 and size per number of probes <10,000.

Kooperative Gesundheitsforschung in der Region Augsburg (KORA) F3 and F4. Genotyping for KORA F3 was performed using the Affymetrix 500K array set, consisting of two chips (StyI and NspI). The KORA F4 samples were genotyped with the Affymetrix human SNP array 6.0. For both studies, genomic DNA from blood samples was used for analysis. Hybridization of genomic DNA was done in accordance with the manufacturer's standard recommendations. Genotyping was

done in the Genome Analysis Centre of the Helmholtz Centre Munich. Genotypes were determined using BRLMM clustering algorithm (Affymetrix 500K array set) and Birdseed2 clustering algorithm (Affymetrix array 6.0). For quality control purposes, we applied a positive control and a negative control DNA every 48 samples (KORA F3) or 96 samples (KORA F4). On the chip level, only subjects with overall genotyping efficiencies of at least 93% were included. In addition, the called gender had to agree with the gender in the KORA study database. After exclusions, 1,644 individuals remained in KORA F3 and 1,814 in KORA F4 for further analysis.

MLPA analysis. We used MLPA to determine changes in the copy number of a region of about 2 Mb on chromosome 16p11.2. Briefly, we designed, using hg18, nine probes within the targeted region, one control probe outside the rearranged region and seven control probes targeting unique position in the genome (Supplementary Table 10). Assays were performed with MRC-Holland reagents according to the manufacturer's protocol[43]. The analysis of the amplification products was performed by capillary electrophoresis in the DNA Analyser 3730XL and using the GeneMapper software v3.7 (Applied Biosystems). The calculations were performed independently for each experiment: we first normalized the MLPA data to minimize the amount of experimental variation, summing all signal values of each control probe for each sample, and then dividing each signal value of each sample by the sum. The normalized signal values were compared to signal values from all other samples in the same experiment, dividing the normalized signal values by the average calculated from all the samples in the same experiment. The product of this calculation is termed dosage quotient (DQ). A DQ value of less than 0.65 or more than 1.25 was considered as copy-number loss or gain, respectively, as previously described[44–46].

Custom array-CGH for the short arm of chromosome 16. DNA samples were labelled with Cy3 and cohybridized to custom-made Nimblegen arrays with Cy5-labelled DNA from the CEPH cell line GM12042. These arrays contained 71,000 probes spread across the short arm of chromosome 16 from 22.0 Mb to 32.7 Mb (at a median space of 45 bp between 27.5 Mb and 31.0 Mb), and 1,000 control probes situated in invariable regions of the X chromosome. DNA labelling, hybridization and washing were performed according to Nimblegen protocols. Scanning was performed using an Agilent G2565BA microarray scanner. Image processing, quality control and data extraction were performed using the Nimblescan software v.2.5.

Defining underweight. Underweight was defined throughout the study as BMI ≤ 18.5 kg per m^2 in adults and ≤ −2 s.d. in children[30,47,48].

Weight, height, BMI and head circumference Z-scores as a function of age. For paediatric cases, weight, height, BMI and head circumference Z-scores were determined for paediatric cases (0–18 years of age) using clinical growth charts specific to the country of origin. Children were ascertained from nine different countries. If charts were only available in percentiles, those measures were transformed into Z-scores using gender-, age- and geographically-matched reference populations (see Statistics).

For the USA and Canada, data from the Center for Disease Control and National Center for Health Statistics (CDC/NCHS) were used to calculate Z-scores[49].

For the French paediatric population, we used French national growth charts[50,51]. For the Swiss paediatric population, we used Swiss national growth charts[52]. For Dutch participants, Dutch national growth charts were used[53]. For Italian, German, Finnish and Austrian cases ($n = 6$), height, weight and BMI Z-scores were estimated using WHO growth charts[54].

To check for discrepancies generated by the use of different growth charts, height, weight and BMI Z-scores were recalculated using WHO growth charts for all cases under five years of age, regardless of origin (http://www.who.int/childgrowth/standards/en/54). Z-scores obtained using the WHO data were not significantly different. These growth standards, developed by the World Health Organization multicentre growth reference study, describe normal child growth from birth to 5 years under optimal environmental conditions. These standards can be applied to all children everywhere, regardless of ethnicity, socioeconomic status and type of feeding[55,56].

If necessary, percentile values were transformed to Z-scores by the inverse-normal density function. When growth charts were unavailable, we used reported LMS parameters (median (M), generalized coefficient of variation (S) and skewness (L)) to obtain Z-scores via the formula:

$$Z\text{-score} = \begin{cases} \frac{(X/M)^L - 1}{L*S}, & L > 0 \\ \ln(X/M)/S, & L = 0 \end{cases},$$

in which X is the observed value.

In adults (>18 years of age), we estimated LMS parameters when these were unavailable from the available sex-, age- and origin-matched Swiss (CoLaus),

Estonian or French control populations. For cases identified from population-based cohorts, Z-scores were directly inferred from the cohort.

Gene expression. We established lymphoblastoid cell lines from deletion and duplication carriers, as well as from controls (Supplementary Table 12), by transforming peripheral blood mononuclear cells with Epstein–Barr virus. Patients and controls were enrolled after obtaining appropriate informed consent via the physicians in charge, and approval by the ethics committee of the University of Lausanne. More control cell lines were obtained from Coriell Institute for Medical Research (http://www.coriell.org/) (Supplementary Table 12). SYBR Green real-time quantitative PCR (RT–PCR) was performed as previously described[22,57]. Briefly, 1 μg of total RNA from lymphoblastoid cell lines was converted to complementary DNA using Superscript VILO (Invitrogen) primed with a mixture of oligo(dT) and random hexamers. Oligos were designed using the PrimerExpress program (Applied Biosystems) with default parameters (Supplementary Table 11). Non-intron-spanning assays were tested for genomic contamination in standard ± reverse transcriptase reactions. The amplification efficiency of each primer pair was tested in a cDNA dilution series, as previously described[58]. A full list of genes mapping in the rearranged interval, and exclusion criteria, are presented in Supplementary Table 1. All RT–PCR reactions were performed in a 10-μl final volume and triplicates per sample. The setup in a 384-well plate format was performed using a Freedom EVO robot (TECAN) and assays were run in an ABI 7900 sequence detection system (Applied Biosystems) with the following amplification conditions: 50 °C for 2 min, 95 °C for 10 min, and 45 cycles of 95 °C 15 s, then 60 °C for 1 min. A final incubation of 95 °C for 15 s followed by 60 °C for 15 s was carried out to establish a dissociation curve. Each plate included the appropriate normalization genes to control for any variability between plate runs. Raw threshold cycles (Ct) values were obtained using SDS2.4 (Applied Biosystems). To calculate the normalized relative expression ratio of individuals carrying the CNV and of controls, we used Biogazelle qBase Plus software[59] including geNorm[60]. This program identified appropriate normalization genes (*EEF1A1*, *RPL13*, *GUSB* and *TBP*) having a gene-stability measure of M = 0.25. We note that one gene, *LAT*, showed a very high expression profile in one of the duplication samples (DASYL, Supplementary Table 13), reaching a relative expression value of 27.3 (s.e.m. = 1.37), compared to an average expression for other duplications of 1.89 (s.e.m. = 0.51). We cannot exclude that this finding is genuine (and confirmed it in a second experiment), but it was removed from further analyses as an outlier to give a more accurate overview of expression profiles for these genes.

In silico analysis was performed to check for brain, and specifically hypothalamus, expression of genes in the rearranged 16p11.2 interval (Supplementary Table 1). This was done using Allen Brain Atlas Resources, available from http://www.brain-map.org.

Cases with major neurological signs. Major neurological signs were defined by moderate to severe hypotonia, hypertonia, ataxia, spasticity, hyperreflexia, hyporeflexia and/or extra-pyramidal signs, and by the presence of epilepsy.

Statistics. Student's *t*-test: one-tailed *t*-tests were performed to test whether duplication carriers have Z-score values lower than zero for BMI, height and weight. We found this analysis more suitable than linear regression analysis, correcting for confounding factors such as sex and age, because these anthropometric traits have a highly nonlinear dependence on these factors, as can be observed in control populations.

Kruskal–Wallis test: this was used to test differences in the gene expression pattern between deletion and duplication carriers and control individuals. Because expression values are not necessarily normally distributed, this test is more adequate than a classical one-way analysis of variance. To test pairwise differences, we computed the difference in mean group rank with its 95% confidence interval (as provided by the multcompare function in Matlab). Correction for multiple testing was done using a Bonferroni adjustment.

Multiple testing: we determined false-discovery-rate-based thresholds for association *P*-values for each phenotype, to correct for multiple testing. For each phenotype, we replaced the observed Z-scores with numbers randomly drawn from a standard normal distribution and performed the same *t*-tests for the same strata. The procedure was repeated 1,000 times. For various *P*-value thresholds, we asked how many tests would be declared significant for the null set on average (over the 1,000 random draws). The false discovery rate was estimated as the ratio of this number and the actual number obtained for the observed Z-scores. Thus, we controlled the dependence between nested tests.

Relative risk: among adults, we defined underweight as a BMI <18.5 (WHO criteria). The estimated relative risk is the ratio of the fraction of underweight individuals among duplication carriers versus our control group. The standard error of log(relative risk) and its significance were calculated as previously described[61]. In our control group (population-based cohorts), the frequency of

being underweight is 1.9% (38 males and 148 females out of 9,470). Owing to the fact that being underweight decreases with age in the general population, we resampled our control group to ensure precise age-matching.

31. Coin, L. J. *et al.* cnvHap: an integrative population and haplotype-based multiplatform model of SNPs and CNVs. *Nature Methods* **7**, 541–546 (2010).
32. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
33. Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657–663 (2007).
34. Colella, S. *et al.* QuantiSNP: an objective Bayes Hidden-Markov model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res.* **35**, 2013–2025 (2007).
35. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665–1674 (2007).
36. Korn, J. M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature Genet.* **40**, 1253–1260 (2008).
37. Marioni, J. C. *et al.* Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. *Genome Biol.* **8**, R228 (2007).
38. Diskin, S. J. *et al.* Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.* **36**, e126 (2008).
39. Bengtsson, H., Irizarry, R., Carvalho, B. & Speed, T. P. Estimation and assessment of raw copy numbers at the single locus level. *Bioinformatics* **24**, 759–767 (2008).
40. Bengtsson, H., Ray, A., Spellman, P. & Speed, T. P. A single-sample method for normalizing and combining full-resolution copy numbers from multiple platforms, labs and analysis methods. *Bioinformatics* **25**, 861–867 (2009).
41. Huang, J. *et al.* Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum. Genomics* **1**, 287–299 (2004).
42. Wang, B. *et al.* Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science* **316**, 1194–1198 (2007).
43. Schouten, J. P. *et al.* Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30**, e57 (2002).
44. Bunyan, D. J. *et al.* Dosage analysis of cancer predisposition genes by multiplex ligation-dependent probe amplification. *Br. J. Cancer* **91**, 1155–1159 (2004).
45. Fernández, L. *et al.* Comparative study of three diagnostic approaches (FISH, STRs and MLPA) in 30 patients with 22q11.2 deletion syndrome. *Clin. Genet.* **68**, 373–378 (2005).
46. Slater, H. R. *et al.* Rapid, high throughput prenatal detection of aneuploidy using a novel quantitative method (MLPA). *J. Med. Genet.* **40**, 907–912 (2003).
47. Mei, Z. *et al.* Validity of body mass index compared with other body-composition screening indexes for the assessment of body fatness in children and adolescents. *Am. J. Clin. Nutr.* **75**, 978–985 (2002).
48. Physical. status: the use and interpretation of anthropometry. Report of a WHO expert committee. *World Health Organ. Tech. Rep. Ser.* **854**, 1–452 (1995).
49. Kuczmarski, R. J. *et al.* CDC growth charts: United States. *Adv. Data* **314**, 1–27 (2000).
50. Sempé, M., Pedron, G. & Roy-Pernot, M. P. *Auxologie, Méthode et Séquences.* (Théraplix, 1979).
51. Rolland-Cachera, M. F. *et al.* Body mass index variations: centiles from birth to 87 years. *Eur. J. Clin. Nutr.* **45**, 13–21 (1991).
52. Prader, A., Largo, R. H., Molinari, L. & Issler, C. Physical growth of Swiss children from birth to 20 years of age. First Zurich longitudinal study of growth and development. *Helv. Paediatr. Acta.*, Suppl. 52, 1–125 (1989).
53. Fredriks, M. *Growth Diagrams, 1997. Fourth Dutch Nation-wide Survey.* 233–242 (Bohn Stafleu Van Loghum, 1997).
54. de Onis, M. *et al.* Development of a WHO growth reference for school-aged children and adolescents. *Bull. World Health Organ.* **85**, 660–667 (2007).
55. de Onis, M., Garza, C., Onyango, A. W. & Borghi, E. Comparison of the WHO child growth standards and the CDC 2000 growth charts. *J. Nutr.* **137**, 144–148 (2007).
56. Mei, Z., Ogden, C. L., Flegal, K. M. & Grummer-Strawn, L. M. Comparison of the prevalence of shortness, underweight, and overweight among US children aged 0 to 59 months by using the CDC 2000 and the WHO 2006 growth charts. *J. Pediatr.* **153**, 622–628 (2008).
57. Molina, J. *et al.* Abnormal social behaviors and altered gene expression rates in a mouse model for Potocki–Lupski syndrome. *Hum. Mol. Genet.* **17**, 2486–2495 (2008).
58. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25**, 402–408 (2001).
59. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8**, R19 (2007).
60. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, 1–11 (2002).
61. Morris, J. A. & Gardner, M. J. Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *Br. Med. J. (Clin. Res. Ed.)* **296**, 1313–1316 (1988).

ARTICLE

Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*

Xiangchao Gan[1]*, Oliver Stegle[2]*, Jonas Behr[3]*, Joshua G. Steffen[4]*, Philipp Drewe[3]*, Katie L. Hildebrand[5], Rune Lyngsoe[6], Sebastian J. Schultheiss[3], Edward J. Osborne[4], Vipin T. Sreedharan[3], André Kahles[3], Regina Bohnert[3], Géraldine Jean[3], Paul Derwent[7], Paul Kersey[7], Eric J. Belfield[8], Nicholas P. Harberd[8], Eric Kemen[9], Christopher Toomajian[5], Paula X. Kover[10], Richard M. Clark[4], Gunnar Rätsch[3] & Richard Mott[1]

Genetic differences between *Arabidopsis thaliana* accessions underlie the plant's extensive phenotypic variation, and until now these have been interpreted largely in the context of the annotated reference accession Col-0. Here we report the sequencing, assembly and annotation of the genomes of 18 natural *A. thaliana* accessions, and their transcriptomes. When assessed on the basis of the reference annotation, one-third of protein-coding genes are predicted to be disrupted in at least one accession. However, re-annotation of each genome revealed that alternative gene models often restore coding potential. Gene expression in seedlings differed for nearly half of expressed genes and was frequently associated with *cis* variants within 5 kilobases, as were intron retention alternative splicing events. Sequence and expression variation is most pronounced in genes that respond to the biotic environment. Our data further promote evolutionary and functional studies in *A. thaliana*, especially the MAGIC genetic reference population descended from these accessions.

Interpreting the consequences of genetic variation has typically relied on a reference sequence, relative to which genes and variants are annotated. However, this may cause bias, because genes may be inactive in the reference but expressed in the population[1], suggesting that sequencing and re-annotating individual genomes is necessary. Advances in sequencing[2] make this tractable for *Arabidopsis thaliana*[3–5], whose natural accessions (strains) are typically homozygous. Relative to the 119-megabase (Mb) high-quality reference sequence from Col-0 (ref. 6), diverse accessions harbour a single nucleotide polymorphism (SNP) about every 200 base pairs (bp) (ref. 3), and indel variation is pervasive[3,7,8]. Characterizing this variation is crucial for dissecting the genetic architecture of traits by quantitative trait locus mapping in recombinant inbred lines (see, for example, ref. 9) or genome-wide association in natural accessions[10].

Here we have sequenced and accurately assembled the single-copy genomes of 18 accessions that, with Col-0, are the parents of more than 700 Multiparent Advanced Generation Inter-Cross (MAGIC) lines[9], similar to the maize Nested Association Mapping (NAM)[11] population and the murine Collaborative Cross[12]. These accessions comprise a geographically and phenotypically diverse sample across the species[9]. Using the genomes, seedling transcriptomes and computational gene predictions we have characterized the ancestry, polymorphism, gene content and expression profile of the accessions. We show that the functional consequences of polymorphisms are often difficult to interpret in the absence of gene re-annotation and full sequence data. The assembled genomes also contribute to the *A. thaliana* 1001 Genomes Project[3–5,13].

Genome sequencing, assembly and variants

We assembled the 18 genomes so that single-copy loci would be contiguous, with less than one assembly error per gene, and therefore

suitable for annotation. Accessions were sequenced with Illumina paired-end reads[2] (Supplementary Table 1), generally with two libraries with 200-bp and 400-bp inserts and reads of 36 and 51 bp, respectively, to between 27-fold and 60-fold coverage. Each genome was assembled by using five cycles of iterative read mapping[14] combined with *de novo* assembly[15] (Supplementary Information sections 2 and 3, and Supplementary Tables 1 and 2). We aligned reads to the final assemblies to detect polymorphic regions[8] lacking read coverage (2.1–3.7 Mb per accession; Supplementary Table 3 and Supplementary Fig. 2). At unique loci, polymorphic regions probably reflect complex polymorphisms[3,8]. The average N50 length (the contig size such that 50% of the entire assembly is contained in contigs equal to or longer than this value) of contiguous read coverage between polymorphic regions was 80.8 kb (Supplementary Table 4).

To report complex alleles consistently, we defined all variants against the multiple alignment consensus of Col-0 and the assembled genomes. For each accession there were 497,668–789,187 single-base differences from Col-0, and about 45,000 ambiguous nucleotides (Supplementary Table 5). The latter may reflect heterozygosity (particularly in Po-0; Supplementary Figs 5–7) or copy-number variants, and they were largely in transposable elements and repeats covering 21.9% of the genome (Supplementary Information section 5.1, and Supplementary Figs 8 and 9). Of 3.07 million SNPs, 45.2% were private to single accessions.

We identified 1.20 million indels, and 104,090 imbalanced substitutions, in which a sequence in Col-0 was replaced by a different sequence (Supplementary Tables 3 and 7). Although 57.5% of indels or imbalanced substitutions were shorter than 6 bp, 1.9% were longer than 100 bp, and overall 14.9 Mb of Col-0 sequence was absent in one or more accessions (Fig. 1a and Supplementary Fig. 8). The assemblies were about 1.6% and about 4.3% shorter than the reference (including

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [2]Max Planck Institute for Intelligent Systems and Max Planck Institute for Developmental Biology, Spemannstraße 38, 72076 Tübingen, Germany. [3]Friedrich Miescher Laboratory, Max Planck Society, Spemannstraße 39, 72076 Tübingen, Germany. [4]Department of Biology, University of Utah, Salt Lake City, Utah 84112-5330, USA. [5]Department of Plant Pathology, Kansas State University, Manhattan, Kansas 66506-5502, USA. [6]Department of Statistics, University of Oxford, South Parks Road, Oxford OX1 3TG, UK. [7]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [8]Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK. [9]The Sainsbury Laboratory, Norwich NR4 7UH, UK. [10]Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK.
*These authors contributed equally to this work.
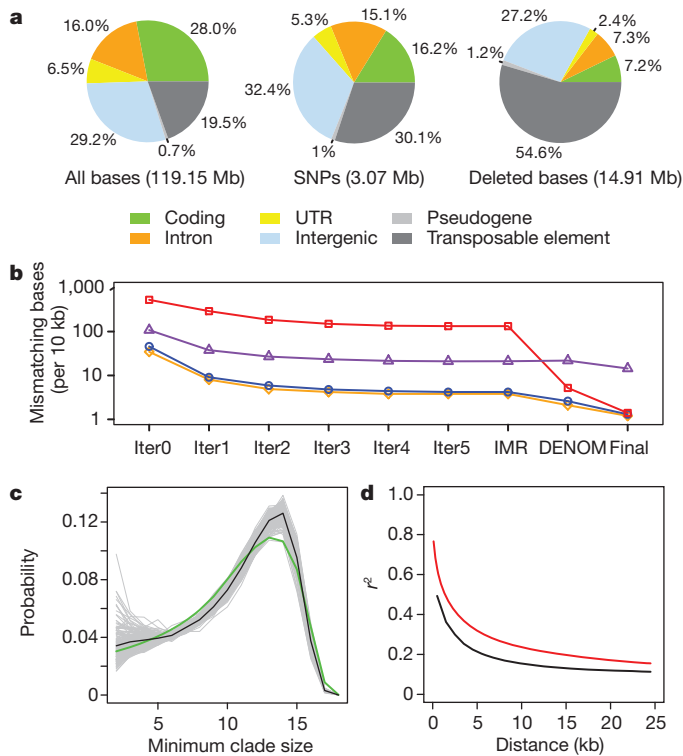
Figure 1 | Assembly and variation of 18 genomes of *A. thaliana*.
a, Classification of sequence, SNPs and indels based on the Col-0 genome.
b, Assembly accuracy (*y* axis; base substitution errors per 10 kb) measured relative to four validation data sets at each of eight stages in the IMR/DENOM assembly pipeline (*x* axis). Bur-0 survey (blue line): 1,442 survey sequences (about 417 bp each) in predominantly genic regions[19]; Bur-0 divergent (red line): 188 sequences (each about 254 bp) highly divergent from Col-0 (ref. 3); Ler-0 nonrepetitive (orange line): a predominantly single-copy 175-kb Ler-0 sequence on chromosome 5; Ler-0 repetitive (purple line): a highly repetitive 339-kb Ler-0 locus on chromosome 3 (ref. 18; Supplementary Information section 4). Iter, iteration. **c**, Genome-wide distribution of the minimum clade size for all pairs of accessions (excluding Po-0). Each pair is represented by a grey line, the mean over all pairs by the black line and the random distribution by the green line. **d**, Decay in linkage disequilibrium with distance (Po-0 excluded). The black line shows r^2 between SNPs; the red line shows phylogenetic r^2 (Supplementary Information section 6).

and excluding polymorphic regions, respectively), probably reflecting limitations in detecting long insertions. Although sequence differences were enriched in transposable-element and intergenic regions, about 17% of bases deleted in one or more accessions were annotated as genic in Col-0 (Fig. 1a and Supplementary Fig. 8). The density of sequence differences is greater than between classical inbred strains of mice[16], but less than between lines of maize[17].

Both iterative and *de novo* assembly improved accuracy, with the latter being most effective at divergent loci (Fig. 1b, Supplementary Table 2 and Supplementary Fig. 10). As assessed with about 1.2 Mb of genomic dideoxy data[3,18,19] (Supplementary Information section 4), the substitution error rate was about 1 per 10 kb in single-copy regions, and about tenfold higher in transposable-element-rich regions. Further, RNA-seq reads covered about 100,000 SNPs per accession with 99.72% concordance (Supplementary Table 5), and junction sequences for 66 of 68 (97%) long indels and imbalanced substitutions were confirmed by PCR and dideoxy sequencing (Supplementary Table 8). The substitution error rate for our assemblies was comparable to that reported for four other *A. thaliana* genome assemblies[4].

Genome-wide patterns of ancestry

The ancestral relationships of the accessions vary genome-wide. We computed phylogenies[20] across 1.25 million biallelic, non-private SNPs

(Supplementary Information section 6). The ancestry of each pair of accessions within a phylogeny was quantified by using the genome-wide distribution of the minimum clade size of the subphylogeny containing the pair (Fig. 1c). Despite their wide geographical origins, with the exception of Po-0 and Oy-0, all pairs have distributions similar to that of an unstructured sample. The probability of recent co-ancestry is slightly higher than expected for a few pairs of accessions, with extended haplotype sharing at a minority of loci (Supplementary Figs 11–15), perhaps reflecting selective sweeps[7]. Both linkage disequilibrium and correlation between neighbouring phylogenies decrease by 50% within 5 kb (Fig. 1d and Supplementary Fig. 16). Variation among the 18 accessions is similar to a diverse global *A. thaliana* sample[7,8] in nucleotide diversity (Supplementary Figs 11–15), correlation with genomic features (Supplementary Tables 9–12) and structural variants (Supplementary Fig. 17).

Gene annotation and transcript diversity

A naive projection of the coordinates of the 27,206 nuclear protein-coding genes from Col-0 (TAIR10 annotation) onto the 18 genomes predicted that 93.4% of proteins were changed in at least one accession, with 32% of the total being affected by genic deletions, premature termination codons, or other disruptions (Supplementary Table 13). This large tally of disrupted genes implies that reference annotations cannot be transferred reliably; in fact, re-annotation reveals compensating changes, ensuring that many genes encode apparently functional proteins (Fig. 2a). Thus, in 96.2% of the 8,757 genes affected, the naive annotations were replaced by an alternative gene model in at least one accession (Fig. 2b and Supplementary Fig. 18). We predicted new splice sites in 64% of the 2,572 genes with splice site disruptions (in 696 cases the new sites were within 30 bp of the original ones; see, for example, Fig. 2a). Finally, there was evidence of alternative splicing in 2,106 genes (Supplementary Information sections 10.10–10.13).

For genome annotation and expression analyses (for example Figs 2–4), we generated 78-bp RNA-seq reads from two biological replicates of seedling mRNA (about 9.5 million mapped reads per accession, including Col-0; Supplementary Information section 9, and Supplementary Table 14). We integrated read alignments[21] with sequence-based gene predictions[22] by using mGene.ngs (Supplementary Information sections 9–10.3, and Supplementary Fig. 19). On average, 24,681 coding genes were predicted for each accession. Comparison of Col-0 *de novo* predictions with TAIR10 annotations (Supplementary Table 16) showed that these predictions are more accurate (transcript *F*-score 65.2%) than using the genome sequence (mGene[22], 59.6%) or RNA-seq alignments alone (Cufflinks[23], 37.5%; Supplementary Table 17). Finally, we consolidated the *de novo* annotations by incorporating TAIR10 annotations where applicable (Supplementary Information section 10.4, and Supplementary Fig. 20); novel transcript structures for a known TAIR10 gene were only accepted if each newly predicted intron was confirmed by RNA-seq alignments, or if the reference gene model was severely disrupted.

We found, on average, 42,338 transcripts per accession (excluding Col-0), of which 5.5% (2,316) were novel (Table 1 and Supplementary Table 18). In each accession there were, on average, 319 novel genes (or gene fragments) supported by RNA-seq (Table 1); 717 novel genes were found in total, 496 whose sequence was present in Col-0 but not annotated, and 221 absent from the Col-0 genome but present in the *de novo* assemblies of the accessions. We found protein or expressed sequence tag matches for 74.9% of the new genes, primarily from *A. thaliana*, *A. lyrata* or other Brassicaceae species (Supplementary Information sections 10.8 and 10.9).

For accession Can-0, we generated additional independent higher coverage RNA-seq data from seedling, root and floral bud, which we used to confirm 83.3% of re-annotated introns (read alignment over splice junction) and 59.9% of transcripts (confirmation of every intron, or read coverage of 50% of the transcript for single exon
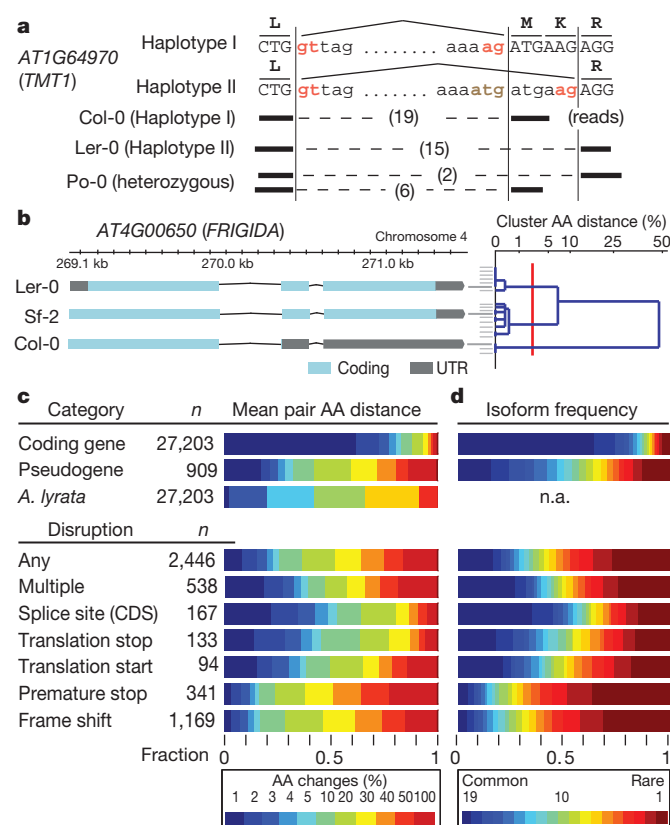
Figure 2 | Transcript and protein variation. a, Example of a splice site change between two haplotypes for the gene *AT1G64970*. Haplotype I (Col-0) is spliced with an intron 6 bp (two amino acids) shorter than haplotype II (Ler-0); Po-0 (heterozygous) shows allele-specific expression of both. **b,** Re-annotation of the *FRIGIDA* locus showing annotations for accessions Sf-2 (functional), and Col-0 (truncated by a premature stop) and Ler-0 (non-functional) (Supplementary Figs 18 and 42). Right: the 19 accessions are shown clustered on the basis of the AA distance between their *FRIGIDA* amino-acid sequences. Common isoform clusters (at distance 2% or less; red line) are shown, leading to three clusters with three, seven and nine accessions. **c,** Proteome diversity for coding genes, pseudogenes and *A. lyrata* genes (top) and for genes with disruptions (bottom). Reported is the fraction of genes with relative AA distance to other accessions (average over pairs) in the given colour-coded interval (Supplementary Information section 10.7). **d,** Frequency of isoforms of coding genes and pseudogenes (top), and those associated with different disruptions (bottom).

transcripts; Table 1). We also obtained additional RNA-seq data for Col-0 and found similar confirmation rates for the reference annotation (Supplementary Table 19). Moreover, for Can-0 we confirmed 72.1% and 84.2% of novel introns and transcripts. Many novel introns stemmed from splice disruptions that tended to be weakly expressed so RNA-seq evidence was scarcer (Supplementary Fig. 22). Finally, more than 75% of novel alternative splicing events were supported by RNA-seq (Supplementary Information section 10.5).

Proteome diversity

To understand the effect of genetic diversity on proteins, it is insufficient to study isolated DNA polymorphisms in the context of the reference annotation. We therefore defined the distance between two amino-acid (AA) sequences by the fraction of amino-acid residues that did not align identically in their global alignment. For example, for FRIGIDA, between Col-0 and Sf-2, a premature stop codon leads to an AA distance of 49% (Fig. 2b). In 77% of proteins, the mean AA distance between all accessions was less than 3% (Fig. 2c). However, on average, 747 proteins per accession had a distance larger than 50% to any TAIR10 protein, with markedly greater variation for pseudogenes. As expected, variation between *A. thaliana* and its congener *A. lyrata*[24] exceeds that observed among *A. thaliana* accessions (Fig. 2c and

Supplementary Fig. 23). Disruptions to splice sites and translation start and stop codons typically caused less severe effects than premature stop codons or frame shifts (Fig. 2c) when compensating splice sites created alternative in-frame splicing (for example Fig. 2a and Supplementary Fig. 24).

Next, we identified protein isoforms across accessions (Fig. 2b, right; distinct isoforms differ by at least roughly 2% AA distance; Supplementary Information section 10.7). For 80% of protein coding genes the most frequent isoform was very common (frequency at least 15 out of 19), whereas isoforms for pseudogenes usually occurred at lower frequency. Moreover, isoforms for large disruptions were rare (frequency 3 or less) for 37% of affected genes (Fig. 2d). This was most pronounced for premature stops and frameshifts, where purifying selection is expected to be strongest.

As expected[3,7], disease resistance genes of the coiled-coil and Toll interleukin 1 receptor subfamilies of the Nucleotide-Binding Leucine Rich Repeat (NB-LRR) gene family were predicted to encode the most variable proteins (Fig. 4a and Supplementary Fig. 26). F-box and defensin-like genes implicated in diverse processes including defence[25,26] were also highly variable. In contrast, housekeeping genes showed little variation.
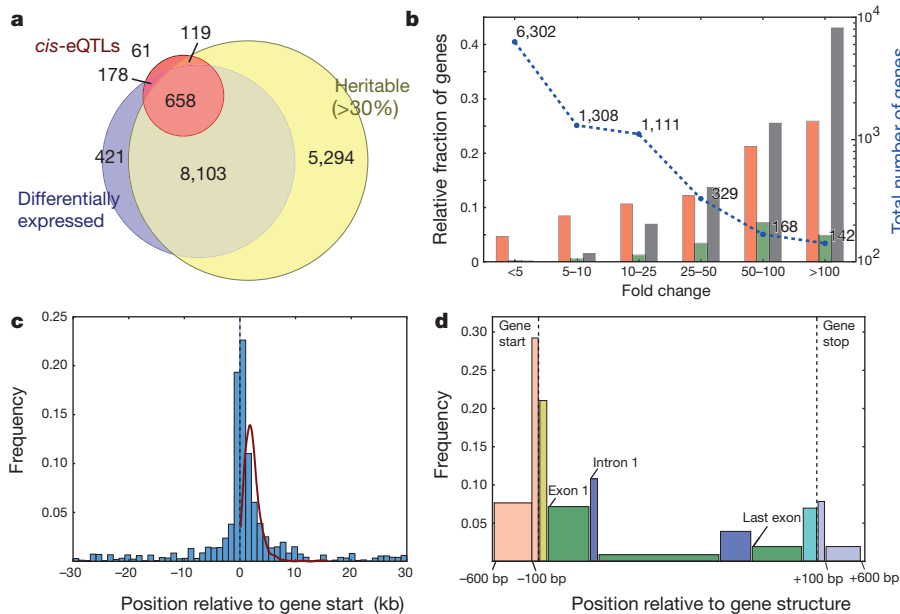
Variation in seedling gene expression

Median expression heritability of protein-coding genes was 39%, similar to that of novel genes (36%) and pseudogenes (38%), and more than for non-coding RNAs (30%) (Supplementary Fig. 27). In total, 75% (20,550) of protein-coding genes (and 21% of non-coding RNAs and 21% of pseudogenes) were expressed in at least one accession (false discovery rate (FDR) 5%), and 46% (9,360) of expressed protein-coding genes were differentially expressed between at least one pair of accessions[27] (Fig. 3a; FDR 5%, Supplementary Information section 11). Of these, 19% (1,750) had more than tenfold expression changes, and 1.5% (142) more than 100-fold (Fig. 3b). For about 60% of genes, at least five accessions contributed to expression variation (Fig. 4d; Supplementary Information section 11.8).

Although the small sample size (19) precludes genome-wide association scans to identify *trans* expression quantitative trait loci (eQTLs), we identified potential *cis*-acting nucleotide variants, copy-number variants and gene structural variants (for example large indels and gene structure changes) associated with expression for 9% (836) of differentially expressed genes (FDR 5%; Supplementary Information section 12.2; we assessed gene-copy-number variation as in Supplementary Information section 12.4). Much of this variation was highly heritable (Fig. 3a). Consistent with identifying likely causal variants, 85% and 93% of associated SNPs and single-nucleotide indels for *cis*-eQTLs were within 5 and 10 kb of the gene, respectively, and were strikingly concentrated in the 100-bp promoter region and 5′ genic sequences (Fig. 3c, d). This was also true for heritable intron retention events, in which most *cis* associations were within the intron or less than 1 kb distant (Supplementary Fig. 32). Our results corroborate the general findings[28–31] of extensive *cis* regulation of gene expression in *A. thaliana*. Neither environmental variation nor population structure markedly affected expression variation (Supplementary Information section 13). Copy-number and structural variants were associated with expression in 3% (240) of differentially expressed genes, including 45% (64 out of 142) of genes with more than 100-fold differences (Fig. 3b), consistent with array studies[29].

Differential gene expression varied by gene ontology (GO) and gene family (Fig. 4b–d, Supplementary Table 24 and Supplementary Figs 39–41). Seventeen of the 18 GO classifications that were enriched for differential expression ($P < 10^{-3}$) concerned response to the biotic environment, including pathogen defence and the production of glucosinolates[32] to deter herbivores (Supplementary Table 24). These include NB-LRR genes (echoing protein variation), of which 74% were differentially expressed at up to 400-fold change, and for which many accessions typically contributed to differential expression

Figure 3 | Quantitative variation of coding gene expression. a, The overlap between heritable (more than 30%) and differentially expressed (FDR 5%) genes, and genes with a *cis*-eQTL (FDR 5%). **b**, Differentially expressed genes and genes with *cis*-eQTLs (FDR 5%) categorized by fold change. Nucleotide variants (orange bars; 647 *cis*-eQTLs) are SNPs and single-base indels; copy-number variants (green bars; 42 *cis*-eQTLs) are regions with elevated coverage in aligned genomic reads in at least one accession; gene structural variants (black bars; 227 *cis*-eQTLs) are accession-specific deletions, insertions or changes to the gene model. **c**, The spatial distribution of nucleotide-variant eQTLs relative to the start of protein-coding genes (FDR 5%, overlapping genes removed; *n* = 647). The line shows density of gene length. **d**, Frequencies of nucleotide-variant eQTLs in protein-coding genes, classified by component (bar widths are proportional to the components' average physical lengths): red bars, upstream; yellow bars, 5′ untranslated region; green bars, coding sequence exons; blue bars, introns; cyan bars, 3′ untranslated region; grey bars, downstream.

(Fig. 4b–d). Patterns for housekeeping genes (such as ribosomal proteins, eukaryotic initiation factors or kinesins) were markedly different: although many were differentially expressed, fold changes were generally small, with variation more often being limited to a few accessions (Fig. 4b–d). Differentially expressed genes generally had much higher nucleotide diversity at synonymous sites relative to other expressed genes, a pattern also observed but less extreme at non-synonymous sites (Supplementary Table 25). This suggests that differences in expression level were not due solely to reduced selective constraint.

The type II MADS box transcription factor family[33] showed striking expression polymorphisms (Fig. 4b–d), including for the *FLOWERING LOCUS C* (*FLC*)[34] and *MADS AFFECTING FLOWERING* (*MAF*) genes[35]. *FLC*, a floral inhibitor expressed highly in accessions that require prolonged cold (vernalization) to flower[36], varied more than 400-fold (Supplementary Fig. 42). F-box and defensin-like genes were exceptional in that expression was restricted in a minority of genes (41% and 12%, respectively; Fig. 4b), perhaps reflecting tissue-specific or environment-specific expression[25,37].

Our data suggest that high turnover for some F-box families in the *A. thaliana* lineage[7] extends to gene expression as well.

Conclusion

Our study goes beyond cataloguing polymorphisms[7,17] to provide genome sequences for a moderately sized population sample (see also refs 4, 16). In doing so, we were able to annotate each genome largely independently of the Col-0 reference. We found that disruptive polymorphisms were frequently compensated for, thereby conserving coding potential and highlighting the limitation of inferring consequences of polymorphisms in the absence of complete sequence data.

Our assemblies are accurate and largely complete in single-copy regions, although additional work will be needed to assemble the roughly 20% of the genome comprising repeats and transposable elements. Disentangling copy variation, long insertions and other genomic rearrangements remains a challenge. The methods we developed are of immediate relevance to the broader *A. thaliana* 1001 Genomes Project[5] and to other organisms, and highlight the importance of RNA-seq data for annotation.
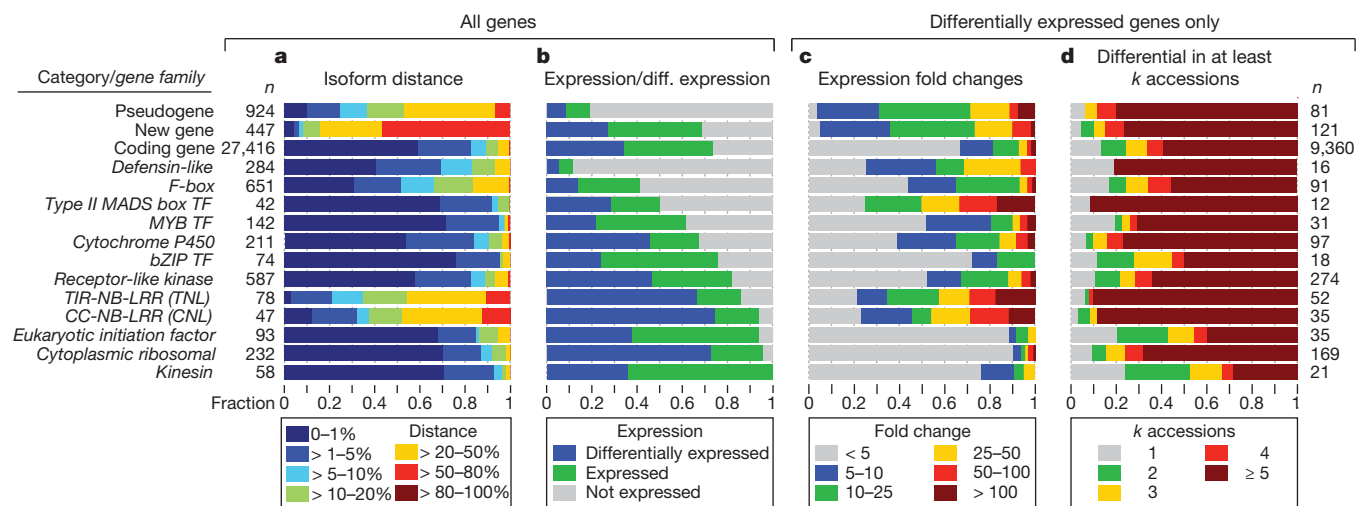
Figure 4 | Protein diversity and gene expression vary by gene category or family. The numbers next to each row are gene counts. The gene families were selected from Supplementary Figs 26 and 39–41 to represent the breadth of observed variation. **a**, Distribution of average AA distances to other accessions (compare with Fig. 2c). **b**, Fraction of unexpressed, expressed and differentially expressed genes (expressed is a superset of differentially expressed). **c**, Distribution of genes categorized by fold change (between lowest and highest across 19 accessions). **d**, Distribution of the numbers of accessions contributing to differential expression. TF, transcription factor; CC, coiled-coil; TIR, Toll interleukin-1 receptor; NB-LRR, nucleotide-binding leucine-rich repeat.

Table 1 | Summary of gene predictions

	Total		Novel	
Type	Per accession	RNA-seq confirmed (%)	Per accession	RNA-seq confirmed (%)
Genes	33,197	62.7	319	88.4
Transcripts	42,338	59.9	2,316	84.2
Introns	127,640	83.3	1,345	72.1
Start codons	33,264	n.a.	503	n.a.
Stop codons	33,720	n.a.	528	n.a.
Intron retentions	1,192	78.1	873	76.5
Exon skips	80	80.5	38	76.7

'Total' and 'novel' are average counts over all 19 accessions. 'RNA-seq confirmed' gives the percentage fully confirmed using independent RNA-seq data (three tissues) for Can-0, the most divergent accession.

Finally, despite using only 19 accessions, we fine-mapped *cis*-eQTLs to small genomic regions (less than 10 kb), suggesting that analogous genome-wide scans in the more than 700 derived MAGIC lines could have single-gene mapping resolution for some loci. Our findings indicate that the MAGIC lines, for which population structure is largely mitigated[9], will be an important and complementary resource to genome-wide association studies in *A. thaliana* populations[10].

METHODS SUMMARY

We used the same seed stocks for Col-0 and the 18 accessions Bur-0, Can-0, Ct-1, Edi-0, Hi-0, Kn-0, Ler-0, Mt-0, No-0, Po-0, Oy-0, Rsch-4, Sf-2, Tsu-0, Wil-2, Ws-0, Wu-0 and Zu-0 that originated the MAGIC lines. DNA and RNA sequencing was performed with standard (DNA) or modified (RNA-seq) Illumina protocols. All methods are described fully in Supplementary Methods; software is available from the authors on request.

1. Johanson, U. *et al.* Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* **290,** 344–347 (2000).
2. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–59 (2008).
3. Ossowski, S. *et al.* Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* **18,** 2024–2033 (2008).
4. Schneeberger, K. *et al.* Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. *Proc. Natl Acad. Sci. USA* **108,** 10249–10254 (2011).
5. Weigel, D. & Mott, R. The 1001 genomes project for *Arabidopsis thaliana. Genome Biol.* **10,** 107, doi:10.1186/gb-2009-10-5-107 (2009).
6. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana. Nature* **408,** 796–815 (2000).
7. Clark, R. M. *et al.* Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana. Science* **317,** 338–342 (2007).
8. Zeller, G. *et al.* Detecting polymorphic regions in *Arabidopsis thaliana* with resequencing microarrays. *Genome Res.* **18,** 918–929 (2008).
9. Kover, P. X. *et al.* A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana. PLoS Genet.* **5,** e1000551 (2009).
10. Atwell, S. *et al.* Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465,** 627–631 (2010).
11. McMullen, M. D. *et al.* Genetic properties of the maize nested association mapping population. *Science* **325,** 737–740 (2009).
12. Durrant, C. *et al.* Collaborative cross mice and their power to map host susceptibility to *Aspergillus fumigatus* infection. *Genome Res.* **21,** 1239–1248 (2011).
13. Cao, J. *et al.* Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genet.* doi:10.1038/ng.911 (28 August 2011).
14. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21,** 936–939 (2011).
15. Li, R. *et al. De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20,** 265–272 (2010).
16. Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* doi:10.1038/nature10413 (in the press).
17. Gore, M. A. *et al.* A first-generation haplotype map of maize. *Science* **326,** 1115–1117 (2009).
18. Lai, A. G., Denton-Giles, M., Mueller-Roeber, B., Schippers, J. H. & Dijkwel, P. P. Positional information resolves structural variations and uncovers an evolutionarily divergent genetic locus in accessions of *Arabidopsis thaliana. Genome Biol. Evol.* advance online publication, doi:10.1093/gbe/evr038 (27 May 2011).
19. Nordborg, M. *et al.* The pattern of polymorphism in *Arabidopsis thaliana. PLoS Biol.* **3,** e196, doi:10.1371/journal.pbio.0030196 (2005).
20. Song, Y. S. & Hein, J. Constructing minimal ancestral recombination graphs. *J. Comput. Biol.* **12,** 147–169 (2005).
21. Jean, G., Kahles, A., Sreedharan, V. T., De Bona, F. & Ratsch, G. in *Current Protocols in Bioinformatics* Ch. 11, Unit 11.6 (Wiley, 2010).
22. Schweikert, G. *et al.* mGene: accurate SVM-based gene finding with an application to nematode genomes. *Genome Res.* **19,** 2133–2143 (2009).
23. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28,** 511–515 (2010).
24. Hu, T. T. *et al.* The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genet.* **43,** 476–481 (2011).
25. Silverstein, K. A., Graham, M. A., Paape, T. D. & VandenBosch, K. A. Genome organization of more than 300 defensin-like genes in *Arabidopsis. Plant Physiol.* **138,** 600–610 (2005).
26. Gagne, J. M., Downes, B. P., Shiu, S. H., Durski, A. M. & Vierstra, R. D. The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis. Proc. Natl Acad. Sci. USA* **99,** 11519–11524 (2002).
27. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106, doi:10.1186/gb-2010-11-10-r106 (2010).
28. Keurentjes, J. J. *et al.* Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. *Proc. Natl Acad. Sci. USA* **104,** 1708–1713 (2007).
29. Plantegenet, S. *et al.* Comprehensive analysis of *Arabidopsis* expression level polymorphisms with simple inheritance. *Mol. Syst. Biol.* **5,** 242, doi:10.1038/msb.2008.79 (2009).
30. West, M. A. *et al.* Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis. Genetics* **175,** 1441–1450 (2007).
31. Zhang, X., Cal, A. J. & Borevitz, J. O. Genetic architecture of regulatory variation in *Arabidopsis thaliana. Genome Res.* **21,** 725–733 (2011).
32. Howe, G. A. & Jander, G. Plant immunity to insect herbivores. *Annu. Rev. Plant Biol.* **59,** 41–66 (2008).
33. Kaufmann, K., Melzer, R. & Theissen, G. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* **347,** 183–198 (2005).
34. Sheldon, C. C. *et al.* The FLF MADS box gene: a repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *Plant Cell* **11,** 445–458 (1999).
35. Ratcliffe, O. J., Kumimoto, R. W., Wong, B. J. & Riechmann, J. L. Analysis of the Arabidopsis *MADS AFFECTING FLOWERING* gene family: *MAF2* prevents vernalization by short periods of cold. *Plant Cell* **15,** 1159–1169 (2003).
36. Lempe, J. *et al.* Diversity of flowering responses in wild *Arabidopsis thaliana* strains. *PLoS Genet.* **1,** 109–118 (2005).
37. Schmid, M. *et al.* A gene expression map of *Arabidopsis thaliana* development. *Nature Genet.* **37,** 501–506 (2005).

Author Contributions R.M., R.C., G.R., P.X.K. and C.T. conceived and led the project. R.M. and X.G. developed and performed genome assembly and analysis. R.C, J.S. and E.O. sequenced the transcriptomes and performed analyses. R.L. performed the ancestral recombination graph analysis. C.T. and K.L.H. performed population genetics analyses. S.J.S., G.J., G.R. and A.K. developed and applied RNA-seq alignment. J.B., G.R., S.J.S., A.K. and R.B. annotated the genomes and analysed their proteins. P.D., G.R. and O.S. performed differential expression analyses. O.S. and P.D. performed genetic association analysis of expression traits. V.T.S. processed data for display in Gbrowse. N.P.H., E.B. and X.G. performed experimental confirmations. P.K. and P.D. constructed the Ensembl version of the variation data. E.K. performed genomic sequencing. R.M., R.C., G.R., P.X.K., C.T., X.G. and O.S. wrote the paper with input from all authors.